Volume 105
Number 1
Spring 2019

# Journal of the WASHINGTON ACADEMY OF SCIENCES




**ISSN 0043-0439** **Issued Quarterly at Washington DC**

# Washington Academy of Sciences

**Founded in 1898**




*Journal of the Washington Academy of Sciences* (ISSN 0043-0439)

Published by the Washington Academy of Sciences
email: wasjournal@washacadsci.org
website: www.washacadsci.org


***The Journal of the Washington Academy of Sciences***

The *Journal* is the official organ of the Academy. It publishes articles on science policy, the history of science, critical reviews, original science research, proceedings of scholarly meetings of its Affiliated Societies, and other items of interest to its members. It is published quarterly. The last issue of the year contains a directory of the current membership of the Academy.

**Subscription Rates**

Members, fellows, and life members in good standing receive the Journal free of charge. Subscriptions are available on a calendar year basis, payable in advance. Payment must be made in US currency at the following rates.

| | |
|---|---|
| US and Canada | $30.00 |
| Other Countries | $35.00 |
| Single Copies (when available) | $15.00 |

**Claims for Missing Issues**

Claims must be received within 65 days of mailing. Claims will not be allowed if non-delivery was the result of failure to notify the Academy of a change of address.

**Notification of Change of Address**

Address changes should be sent promptly to the Academy Office. Notification should contain both old and new addresses and zip codes.

**Postmaster**:

Send address changes to WAS, Rm GL117,
1200 New York Ave. NW
Washington, DC 20005

**Academy Office**

Washington Academy of Sciences
Room GL117
1200 New York Ave. NW
Washington, DC 20005
Phone: (202) 326-8975

Volume 105
Number 1
Spring 2019

# Journal of the WASHINGTON ACADEMY OF SCIENCES




**ISSN 0043-0439** **Issued Quarterly at Washington DC**

# EDITOR'S COMMENTS

Presenting the 2019 Spring issue of the *Journal of the Washington Academy of Sciences*.

I encourage people to write letters to the editor. Please send email (wasjournal@washacadsci.org) comments on papers, suggestions for articles, and ideas for what you would like to see in the Journal. I also encourage student papers and will help the student learn about writing a scientific paper.

First up are two tactile astronomy demos. These are especially useful for students who learn through tactile means. Just how many stars are in the Milky Way? A mere number is difficult to comprehend. This paper addresses that issue.

To follow is a short description of the flu virus and how it can adapt and change. Flu pandemics have killed millions of people. This paper was accepted some months ago when the flu session was in full swing.

Next up is a student paper from Frederick Community College. It discusses the medical uses for garlic and tea tree oil.

Finally a multi-author paper on contextual label smoothing.

The Journal is the official organ of the Academy. Please consider sending in technical papers, review studies, announcements, and book reviews.

We are a peer reviewed journal and need volunteer reviewers. If you would like to be on our reviewer list please send email to the above address and include your specialty.

Sethanne Howard


**Journal of the Washington Academy of Sciences**

**Editor** Sethanne Howard showard@washacadsci.org

## Board of Discipline Editors

The *Journal of the Washington Academy of Sciences* has a twelve member Board of Discipline Editors representing many scientific and technical fields. The members of the Board of Discipline Editors are affiliated with a variety of scientific institutions in the Washington area and beyond — government agencies such as the National Institute of Standards and Technology (NIST); universities such as Georgetown; and professional associations such as the Institute of Electrical and Electronics Engineers (IEEE).

# Tactile Astronomy Demos: Milky Way "Stars like Grains of Sugar" plus Ball and Sun Lunar Phases

**Gene Byrd**

**University of Alabama**

## Abstract

Indoor and outdoor astronomical size/distance demonstrations are well- known. Here we discuss two tactile demos showing *not sizes* but astronomical number and shape. In the first even elementary students appreciate the immense number of Milky Way stars using a 5 lb. bag of fine-grained sugar. Using the approximate size of a grain, a typical bag would be about 1000x1000x1000 grains in length, width and depth thus containing about a billion grains. When the bag is theatrically poured slowly into a container, students can see and, afterward, feel the "multitude" of sugar stars in just one bag. The roughly 100 billion stars in the disk of our Milky Way are comparable to the number of grains in a hundred bags of sugar, far too many to bring to class! Sand can be used if convenient. The second demo dramatically shows the shape and origin of the phases of the Moon.as illuminated by the Sun. Both must be visible on a clear sunny morning or afternoon. Holding a small ball with thumb and forefinger in the Moon's direction magically creates on a "microscopic" scale the same phase for the ball (crescent, half or gibbous) as for the much larger and more distant Moon "beside" the ball.

## Introduction

**INDOOR AND OUTDOOR ASTRONOMICAL** size/distance demonstrations are well-known, *e.g.*, of the huge ratio of the Sun's size versus planets, and the separations of the Sun and planets versus their sizes. The excellent NASA After School Universe program and site: https://imagine.gsfc.nasa.gov/educators/programs/au/ contains exercises along these lines, most notably a paper plate scale model of the Milky Way. Here we discuss a visual and tactile demonstration showing *not sizes* but the "astronomical" number of stars in our Milky Way. We also discuss a tactile demonstration of lunar phases on a micro and macro astronomical scale.

## The Number of Stars in the Milky Way

In the first demonstration we used grains of sugar to help students appreciate the immense number of Milky Way stars. While this concept is probably not totally new, for this author, this demo was triggered by Archimedes' work: *The Sand Reckoner*. With only a few planets and only a few hundred cataloged stars known at that time, Archimedes estimated the number of grains to fill an enlarged universe as necessitated by Aristarchus' heliocentric theory. Today, an immensely larger number of stars in just our Milky Way Galaxy is inferred from modern estimates of the mass of the Galactic disk and bulge.

For an elementary school class, we bought a 5 lb. bag of fine-grained sugar. See Figure 1. The size of a grain is about 0.1 mm so a 10x10x10 cm bag would be about 1000x1000x1000 grains in length, width and depth. Multiplying, together, our bag had about a billion grains. The teacher theatrically poured the bag's grains slowly into a container letting the students see and, afterward, feel the "multitude" of sugar stars from just one bag. There are about 100 billion stars in the disk of our Milky Way. This huge number is comparable to the grains in a hundred bags of sugar. This is far too many to bring to class! Sand can be used if available in a conveniently sized or shaped bag.


Figure 1: A billion grains visually and tactilely displayed

## Moon Phases with a Ball

Again tactilely and visually, the second demonstration dramatically shows the shape and origin of phases of the Moon. For this demonstration, the Sun and Moon must both be visible in a clear sky. The morning sky shortly after sunrise is usually best. The teacher or student has to be alert for good observing conditions and the time a given phase is in the sky. Holding a golf or tennis ball with thumb and forefinger in the Moon's direction magically creates on a "microscopic" scale the same phase for the ball (crescent, half or gibbous) as for the much larger and more distant Moon seen "beside" the ball. See Figure 2 for the arm, ball, Moon, and observer orientation.

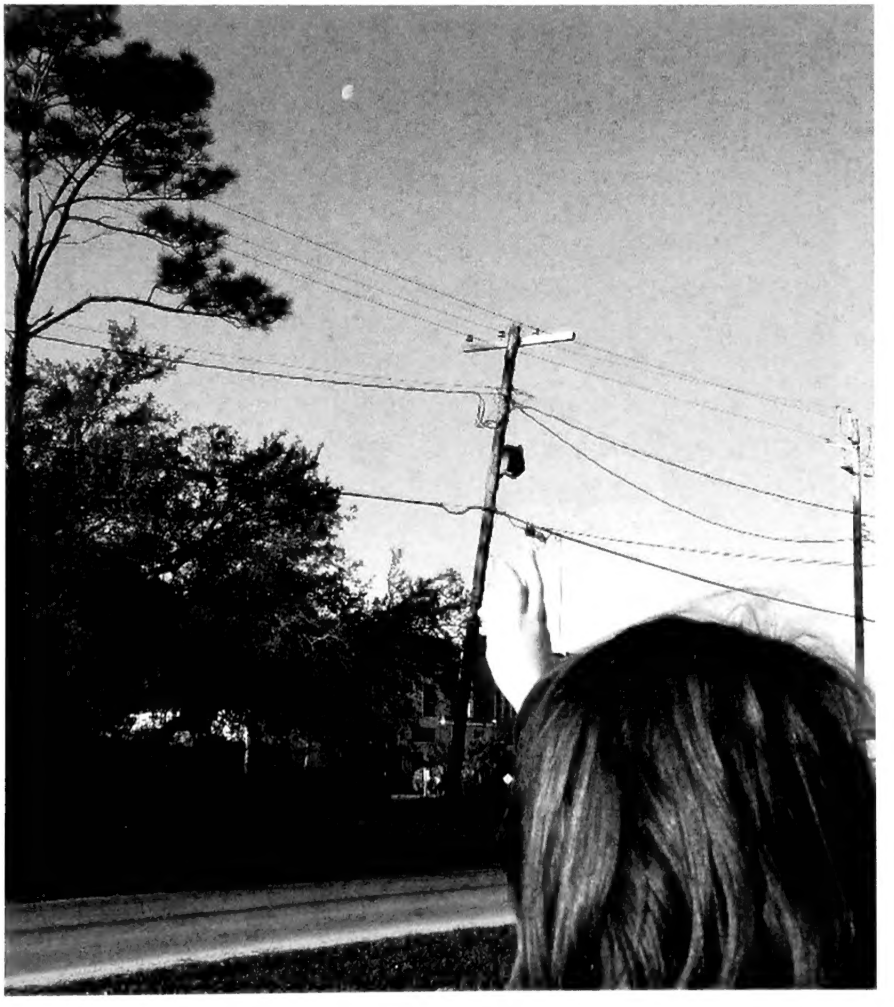


Figure 2: Holding the ball in a line almost between the eye and the sun. This is a clear morning with both the sun and the waning gibbous moon in the sky.

Figure 3 shows a close-up of a golf ball on a push pin in the 3rd quarter position relative to the Sun. If you look carefully, you can see the 3rd quarter moon directly above the golf ball! Note that the "day-night line" terminator orientation matches that of the actual Moon. This is a simple photo taken with a cell phone camera held at the observer's eye. The camera lens must be as close as possible to the eye/golf ball/moon line, not off to one side. The golf ball provides ready-made "craters" which are best seen along the terminator of the ball as on the moon itself through a small telescope or binoculars.

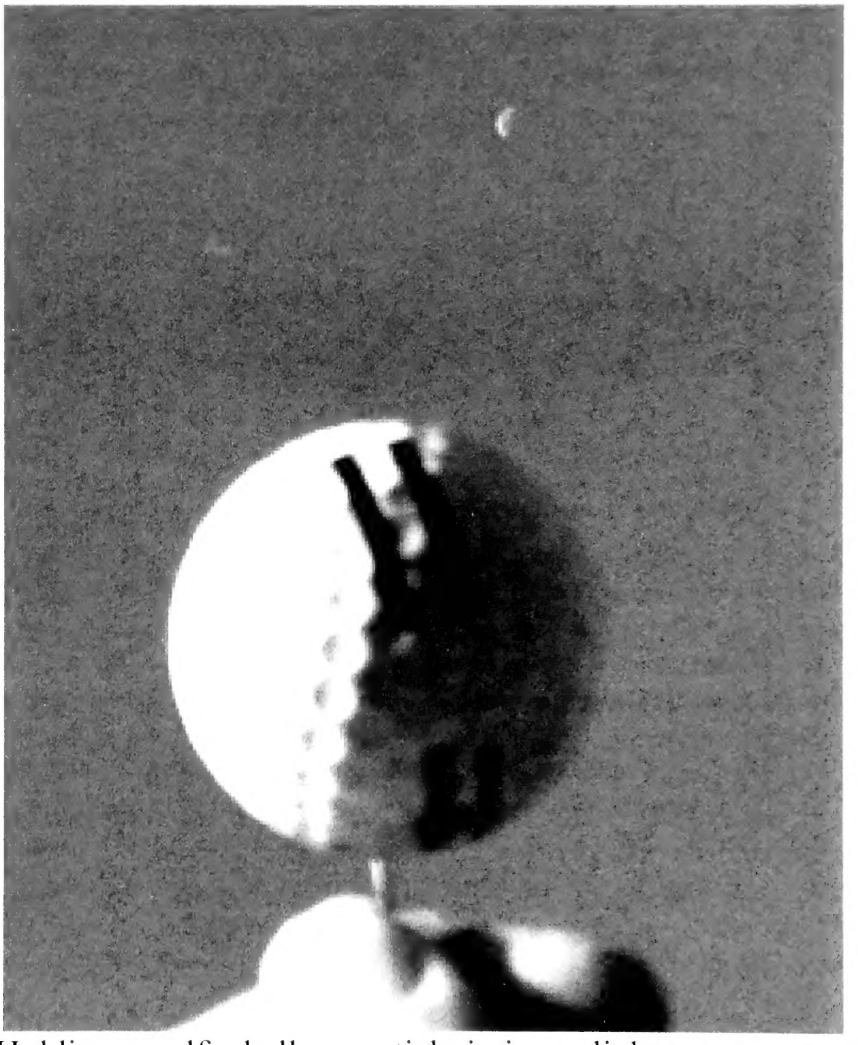


Figure 3: Holding a golf a ball on a stickpin in sunlight to generate phases of the Moon. 3rd quarter is created for both the ball and the Moon is seen above it. The same phase results because of the same Sun-Observer-Moon/ball shape and orientation on a micro and macro size/distance scale.

## Conclusions

We have explored two simple tactile astronomical demonstrations. The first gives a striking visual and tactile "feel" for the billions of stars in the disk of our Milky Way Galaxy. An abstract factor in the Drake Equation for the number of currently existing life and civilizations in our Galaxy is thus made real.

When illuminated by the Sun, we have seen how a hand-held golf ball "magically" shows the same phase as the more distant Moon in the same direction. This provides a strong tactile and visual feel beyond simply looking at a diagram or just using an artificial light and ball alone.

## Acknowledgements

I acknowledge Ms Lavender's Tuscaloosa Capitol School 4th grade and University of Alabama online Astronomy lab course students for serving as test subjects. See https://www.researchgate.net/profile/Gene_Byrd2 for photos *etc*. on this and other educational and research topics

## References

Archimedes' work 287-212 BCE. *The Sand Reckoner*, Ψαμμίτης, http://www.numericana.com/answer/archimedes.htm annotated and translated by Gerard Michon 2002-2015, is a work by Archimedes in which he set out to determine an upper bound for the number of grains of sand that fit into the Universe.

*NASA After School Universe program and site.* https://imagine.gsfc.nasa.gov/educators/programs/au/ Click on *Afterschool Universe Program Leader's Manual* and go to Session 9 on the Milky Way.

Drake Equation. The Drake equation is a probability argument established by Dr. Frank Drake and used to estimate the number of active, communicative extraterrestrial civilizations in the Milky Way galaxy. The equation summarizes the main concepts which scientists must contemplate when considering the question of other radio-communicative life. See https://www.seti.org/drake-equation-index

## Bio

Gene G. Byrd is a Professor Emeritus of Astronomy at the University of Alabama in Tuscaloosa, Alabama.

# The Antigenic Shift or Drift of the Influenza Virus

John J. Paulonis

**Abstract**

Although there is no antigenic shift for this year, the process is quite interesting. I trace the history of the flu and describe antigenic drift and antigenic shift.

**THE FLU WAS FIRST IDENTIFIED** by Hippocrates around 410 BCE, describing a highly contagious illness found in northern Greece. It wasn't until 1357 CE that the term 'influenza' was derived. The word originated from the Italian '*influenza di freddo*' (cold influence) named for an epidemic in Florence, Italy where the people identified that this illness was demonstrated during the colder weather.

The flu was first thought to be a bacterium, but it wasn't until 1931 that a virus in pigs was discovered to be the cause of the flu (in humans, in 1933).

The most infamous pandemic (occurring over a large geographic area, either in a country or the world) was the Spanish Flu of 1918. It has been said that more U.S. soldiers had died from the flu during WWI than from battle itself. (https://www.history.com/topics/inventions/flu )

Influenza viruses have distinct nomenclature depending upon the genetic make-up of the virus. The various particular strains are given nomenclature such as H1N1, more commonly referred to as the "Swine Flu". (The H is an abbreviation for hemagglutinin while the N is an abbreviation for neuraminidase. HA, meaning hemagglutinin antigen, and NA meaning neuraminidase antigen).

We are currently experiencing the 2018 – 2019 Flu Season. In general the influenza virus can undergo a number of changes and may become virulent, even though a person has received an influenza vaccine. This year's influenza activity are listed in Figure 1.

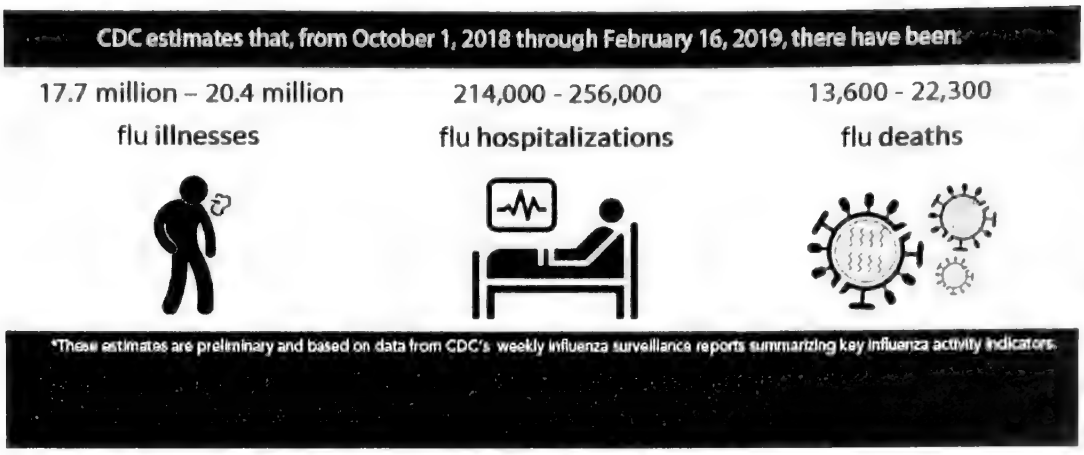



Figure 1 Influenza activity
(Retrieved Feb 2019 from https://www.cdc.gov/flu/index.htm)

According to the CDC, the dominant Influenza A strain which has been predominantly testing positive is (H1N1)pdm09, with one quarter of specimens testing positive for H3N2. Vaccine effectiveness was estimated to be 46% (30%–58%) against illness caused by influenza A(H1N1)pdm09 viruses. (Office of the Associate Director for Communication, Digital Media Branch, Division of Public Affairs. (2019, Feb. 22))

**Antigenic drift** are small changes in the genes of influenza viruses that happen continually over time as the virus replicates. As antigenic changes accumulate, the antibodies created against the older viruses no longer recognize the "newer" virus, and the person can get sick again. See Figure 3.

"**Antigenic shift** is an abrupt, major change in the influenza A viruses, resulting in new hemagglutinin (HA refers to glycoproteins on the surface of influenza viruses which cause red blood cells to agglutinate. The red blood cells clump. HA attaches to cell receptors and initiates the process of virus entry into cells.)[1] and/or new hemagglutinin and neuraminidase (NA)." The function of the NA protein is to remove sialic acid from glycoproteins. It is the cell receptor to which the influenza virus attaches via the HA protein. HA and NA are proteins in influenza viruses

[1] http://www.virology.ws/2013/11/05/the-neuraminidase-of-influenza-virus/

that infect humans. While influenza viruses are changing by antigenic drift all the time, antigenic shift happens only occasionally.[2] See Figure 2.

Avian influenza refers to the disease caused by infection with avian (bird) influenza (flu) Type A viruses. These viruses occur naturally among wild aquatic birds worldwide and can infect domestic poultry and other bird and animal species. Avian flu viruses do not normally infect humans. However, sporadic human infections with avian flu viruses have occurred. (Centers for Disease Control and Prevention, National Center for Immunization and Respiratory Diseases (NCIRD) (2017, Apr. 13))

Such a "shift" occurred in the spring of 2009, when an H1N1 virus with a new combination of genes emerged to infect people and quickly spread, causing a pandemic. When shift happens, most people have little or no protection against the new virus.

**Bio**

J. Paulonis has a Master's of Science in Natural Sciences from the Roswell Park Cancer Institute Graduate Division of the State University of New York at Buffalo and a Master's of International Management from the Thunderbird School of Global Management.

[2] Sep 27, 2017 ( https://www.cdc.gov/flu/about/viruses/change.htm)

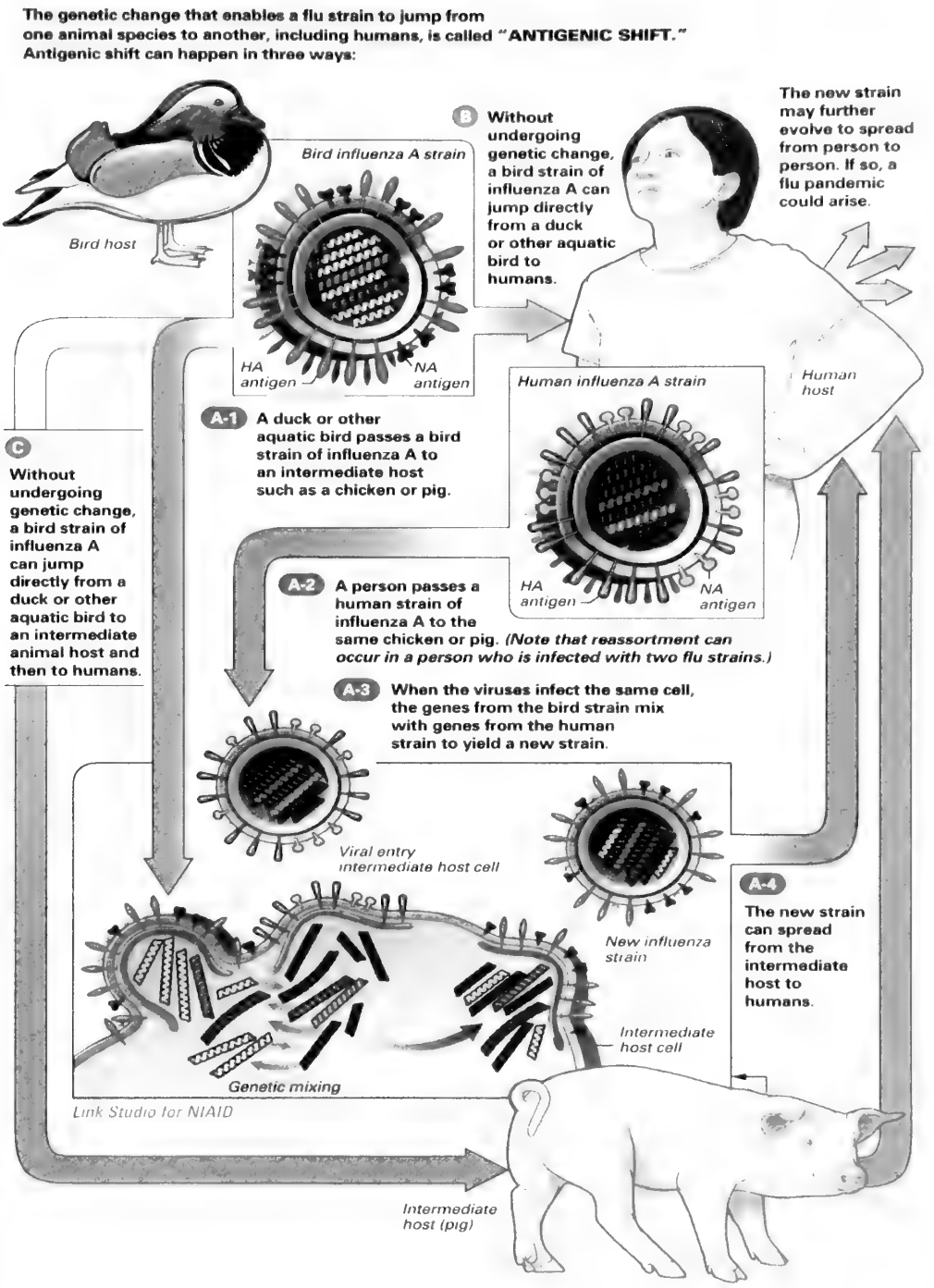
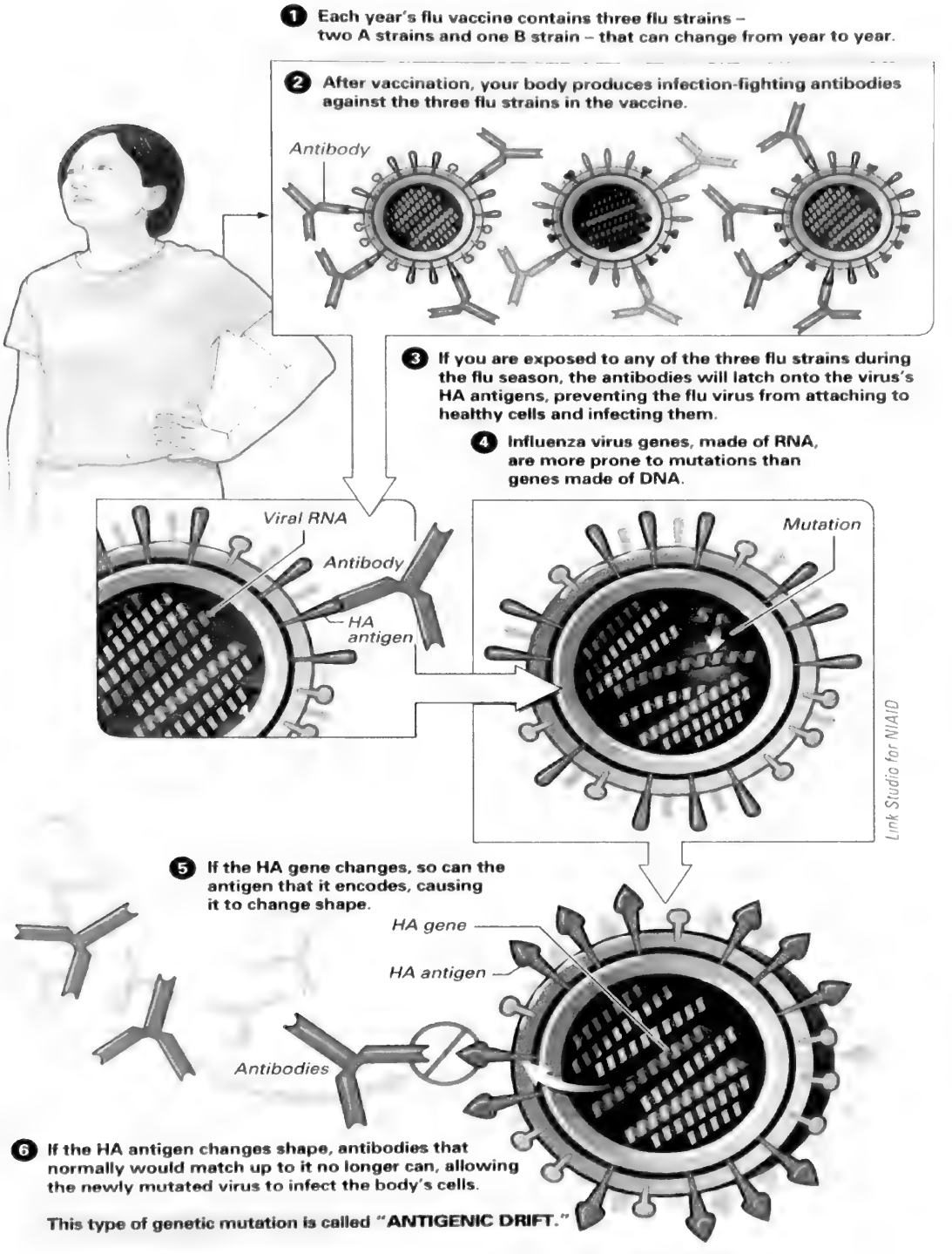



Figure 2 antigenic shift




https://www.verywellhealth.com/what-are-antigenic-drift-and-shift-770400

Figure 3 antigenic drift

## References

History.com Editors. (2018, Aug. 21) *Influenza.* Retrieved from https://www.history.com/topics/inventions/flu

Office of the Associate Director for Communication, Digital Media Branch, Division of Public Affairs. (2019, Feb. 22) *Weekly U.S. Influenza Surveillance Report.* Retrieved from https://www.cdc.gov/flu/weekly/index.htm#whomap

Racaniello, V. (2013, Nov. 5) *The neuraminidase of influenza virus.* Retrieved from http://www.virology.ws/2013/11/05/the-neuraminidase-of-influenza-virus/

Office of the Associate Director for Communication, Digital Media Branch, Division of Public Affairs. (2017, Sept. 27) *How the Flu Virus Can Change: "Drift" and "Shift".* Retrieved from https://www.cdc.gov/flu/about/viruses/change.htm

Duda, K. (2018, Dec. 19) *Antigenic Drift and Shift With the Flu Virus.* Retrieved from https://www.verywellhealth.com/what-are-antigenic-drift-and-shift-770400

Centers for Disease Control and Prevention, National Center for Immunization and Respiratory Diseases (NCIRD) (2017, Apr. 13) *Information on Avian Influenza.* Retrieved from https://www.cdc.gov/flu/avianflu/index.htm

# *In Vitro* Antibacterial Activity of Garlic and Tea Tree Oil

Silvia Godinez, Godfrey Ssenyonga, Judy Staveley

Frederick Community College

## Abstract

To evaluate antibacterial activity of tea tree oil and fresh pure garlic against infectious bacteria preparations of each were combined at different concentrations with cultures of bacteria. The selected essential oil and fresh crushed garlic were screened against one gram-negative bacteria (*Escherichia coli*) and five gram-potentially positive bacteria (*Bacillus cereus, Staphylococcus epidermidis, Bacillus subtilis, and Micrococcus luteus*). Different concentrations (1:1, 1:25, 1:50) were tested using the disc diffusion method. Tea tree essential oil and fresh crushed garlic showed antibacterial activity against one or more bacterial strains. The different concentrations were used to test for differences in antibacterial activity employing the disc diffusion method. The 100% tea tree essential oil and fresh crushed garlic preparations exhibited significant inhibitory effects against the tested bacterial strains. Tea tree oil and the fresh crushed garlic showed promising inhibitory activity even at low concentrations. In conclusion, tea tree oil and crushed fresh garlic showed antibacterial activity against several tested bacterial strains. These findings support the inference that preparations of 100% tea tree oil and of garlic could play a role in inhibiting infection by some gram negative and gram positive bacteria.

## Background

**THE SPREAD OF ANTIMICROBIAL RESISTANT PATHOGENS** is one of the most serious threats to efficacious treatment of microbial diseases. Essential oils and other food plant extracts such as garlic have been used as alternative medical treatments. Many such remedies have been investigated for potentially possible use against a variety of communicable diseases (Zaika, 1988).

Medicinal plants like garlic are used extensively today in food products and in culinary dishes. Fresh garlic has been used for many centuries around the world, especially in the United States, Mexico, Africa, and the Far East. It is scientifically proven that garlic is effectively used against bacterial, viral, mycotic and parasitic infections (Gulsen & Erol, 2010). There is evidence that the garlic plant has immunological properties

that include enhancing the immune system against malignancy and disorders of immune functioning. In this research the potential antibacterial properties of crushed garlic (*Allium sativum*) and its use of antimicrobial potency were investigated against six strains of bacteria. The antibacterial activity was determined using the disc diffusion method.

Essential oils have been shown in many research articles to possess antibacterial, antifungal, antiviral insecticidal and antioxidant properties (Burt, 2004). Tea tree oil has been used for over 100 years as a healing treatment in different countries, particularly for skin conditions. Tea tree oil is best known for its antibacterial activity although it has other likely medicinal properties. To evaluate specifically the antibacterial activity of Tea Tree Oil (*Melaleuca alternifolia*) preparations of different concentrations of the oil were tested against six strains of bacteria. Again the level of antibacterial activity was determined using the disc diffusion method.

## Methods

### *Microorganisms*

Microorganisms were obtained from the Department of Biotechnology, Frederick Community College, Frederick, MD. Six strains of bacteria were used (Table 1). The cultures of bacteria were maintained in their appropriate agar slants at 4°C throughout the study and used as stock cultures. The selected essential oil was screened against one gram-negative bacteria (*Escherichia coli*) and five gram-positive bacteria (*Bacillus cereus, Staphylococcus epidermidis, Serratia marcences, Bacillus subtilis, and Micrococcus luteus*).

The three different concentrations of fresh pure garlic (*Allium sativum*) and Tea Tree Oil (*Melaleuca alternifolia*) (1:1, 1:25, and 1:50) were prepared using the disc diffusion method.

**Table 1**

| 6 Strains of bacteria | Type of bacteria | |
|---|---|---|
| *Bacillus subtilis* | Gram positive | ATCC 6633 |
| *Bacillus cereus* | Gram positive | ATCC 14579 |
| *Staphylococcus epidermis* | Gram positive | ATCC 12228 |
| *Serratia marcescens* | Gram positive | ATCC |
| *Micrococcus luteus* | Gram positive | ATCC 4698 |
| *Escherichia coli* | Gram negative | ATCC 25922 |

***Essential oils***

100% concentration tea tree oil was obtained and was used in this study (Table 2). This essential oil was selected based on previous literature in which it has been used in alternative medical practices and experimentation.

***Fresh Crushed Garlic***

Fresh chopped garlic was obtained from a local grocery store, and used this study (Table 2). This fresh garlic was selected based on previous literature used in alternative medical experiments.

***Antibacterial Assay***

Screening of the tea tree oil and crushed garlic was conducted to estimate antibacterial activity. The antibacterial assay was conducted with the disk diffusion method. This process is normally used as a preliminary check. The antibacterial assay was performed by using a 45 h culture at 37°C incubation. Five hundred microliters of the suspensions were spread over the plates containing BBL nutrient agar using a sterile inoculating loop in order to get a uniform microbial growth on both control and test plates.

The tea tree and fresh crushed garlic were dissolved in an aqueous solution of water and dimethylsulfoxide (DMSO).

**Table 2**

| Essential Oils | Botanical Name | Properties |
|---|---|---|
| Tea Tree Oil | Species: ***M. alternifolia***<br>Kingdom: Plantae<br>Clade: Angiosperms, Eudicots<br>Family: Myrtaceae<br>Genus: Melaleuca | Antiseptic, antibacterial, antiviral, antifungal, and anti-inflammatory agent. |
| Fresh Garlic | Species: ***A. sativum***<br>Kingdom: Plantae<br>Clade: Angiosperms, Monocots<br>Family: Amaryllidaceae<br>Subfamily: Allioideae<br>Genus: Allium | Antiseptic, antibacterial, antiviral, antifungal, and anti-inflammatory agent |

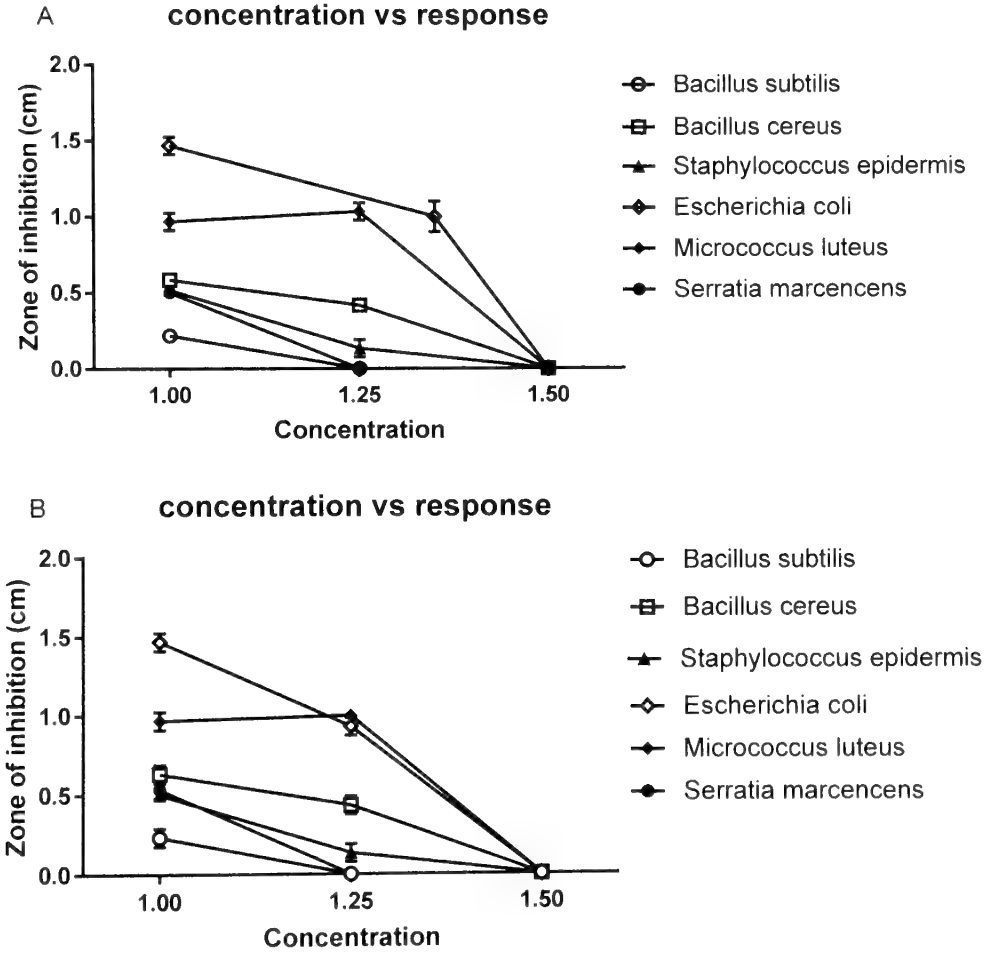
Under aseptic conditions empty sterilized discs (5, 6 mm diameter) were infused with different concentrations (1:1, 1:25, and 1:50) of the respective tea tree oil and fresh crushed garlic. They were placed on the BBL nutrient agar surface. The paper disc was saturated with aqueous concentrations of the tea tree oil and fresh crushed garlic. DMSO was mixed in a microcentrifuge with different concentrations of tea tree oil and fresh garlic. The standard disc was saturated with mixed concentrations and placed on the petri dish. A standard disc containing DMSO was used as reference control for every species of bacterium. All petri dishes were sealed with sterile laboratory tape to avoid evaporation of the test samples. The plates were left for 30 min at room temperature to allow for the diffusion of oil, and then they were incubated at 37°C for 45 h. After the incubation period, the zone of inhibition was measured in centimeters with a caliper and data were recorded. Studies and data were collected over a

series of months.

## Results

### *Antimicrobial activity of Tea Tree oil and garlic oils*

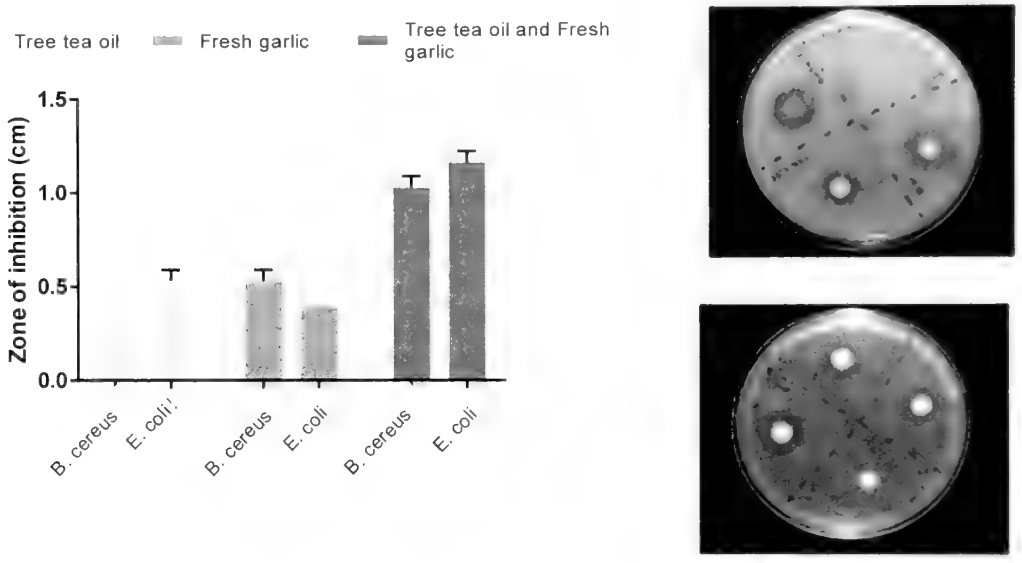
We tested the effects of tea oil and garlic against six types of bacteria in three different concentrations. The Tree Tea Oil showed a greater inhibitory effect on cereus and *E. coli* and the smallest effect was observed on the *S. epidermis* (Graph 1A). While fresh garlic extract was more effectively inhibits *M. luteus* and *E.coli* (Graph 1B), both effects are clearly observed at the high and medium concentrations of essential oils. The values were compared against negative control.




**Graph 1**.Antimicrobial activity graphics. A) Concentration vs response graph of the inhibition from Tree tea oil. B) Concentration vs response graph of the inhibition from Fresh garlic.

### *Tea Tree Oil and Fresh Garlic Extract showed a synergic effect*

We tested the inhibition of both compounds (Tree tea oil and Fresh garlic) at the 1.25 concentration with two bacteria (*Bacillus cereus* and *Escherichia coli)*. The bacteria showed inhibition in the presence of both extracts. The results indicated that the effect of inhibition of these two extracts together was more effective than the activity of each one. Figure 1 shows the plates where the inhibition when the tea and garlic were mixed and Graph 2 shows the corresponding data.




Graph 2 and Figure 1. Synergic effect. Right panel. Graph of inhibition with Tree tea oil, Fresh garlic and mixture. Left panel. Upper plate, inhibition of *B. cereus* by tree tea oil, fresh garlic and mixture. Lower plate, inhibition of *E. coli* by tree tea oil, fresh garlic and mixture.

Dilution 1:25 was shown to have antibacterial activity against *E.coli* and *B. cereus*. The mix of Tea Tree Oil and Fresh Garlic Extract showed a synergistic effect.

### *Preliminary results – antibacterial Allicin Identification*

Several reports have described that the main component of the antibacterial activity of garlic is allicin. Obtaining this biologically active component compound is difficult. We assessed the presence of allicin from a commercial product by analyzing it with Infrared Fourier Transform Spectroscopy. Figure 2 shows the observations and comparisons between

the allicin and fresh garlic extract. The spectra are demonstrated by the main peaks.

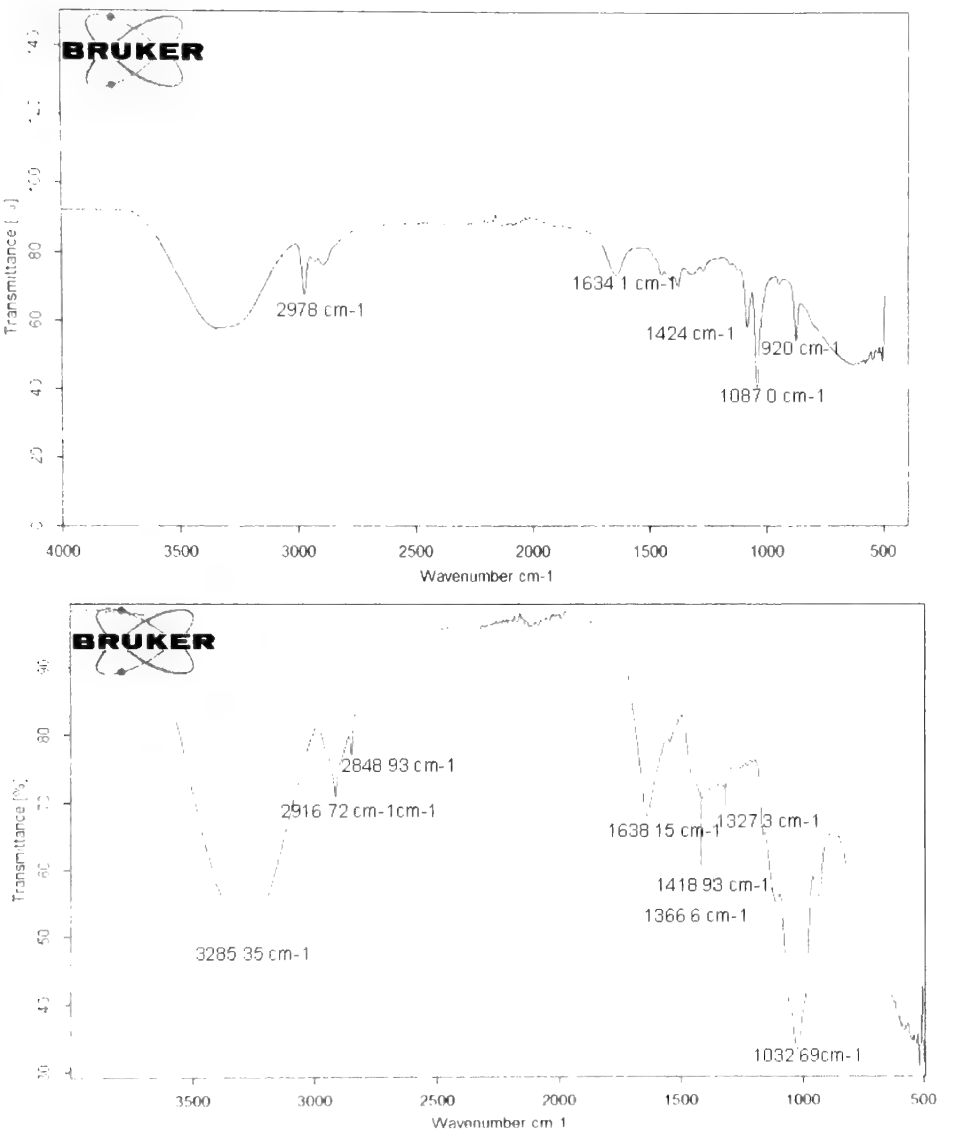
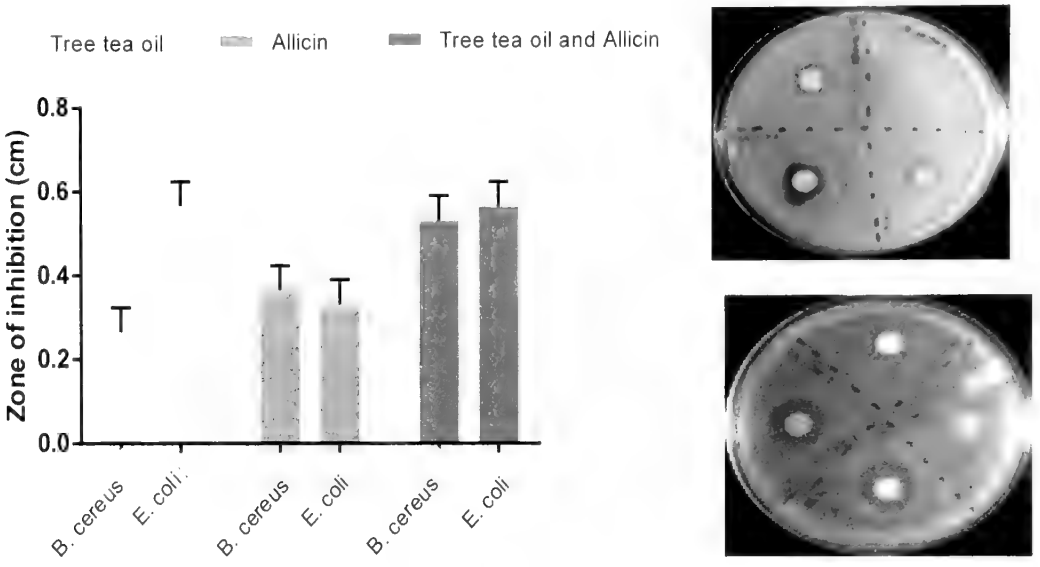



Figure 2. Spectra of the Infrared Fourier Transform. Top panel, Spectrum of allicin from capsules. Bottom panel, spectrum of the bulb extract of the garlic. Peaks identified, 988 $cm^{-1}$ prob. Flex (δ) R-CH=$CH_2$; 1087 $cm^{-1}$ S=O; 1424 $cm^{-1}$ δ CH; 1634. 1 $cm^{-1}$ C=C; -1 ν sim $CH_2$ and ν asim $CH_2$.

***Preliminary Results Antibacterial activity- Allicin)***

The results showed the characteristic peaks of allicin were presented in the fresh garlic extract. We evaluated the activity of the allicin from capsules and used seem conditions by identifying if it had a synergistic effect when mixed the Allicin and Tree tea oil. Figure 3 showed that allicin inhibited the growth of bacteria; however, the inhibition of the mixture did not show a synergistic effect.




Graph 3 and Figure 3. Effects of allicine. Right panel. Graph of inhibition of Tree tea oil, allicin and mixture. Left panel. Upper plate, inhibition of *B. cereus* by tree tea oil, allicine and mixture. Lower plate, inhibition of *E. coli* by tree tea oil, allicin and mixture.

**Conclusion**

The 100% Tea Tree essential oil preparation (*Melaleuca alternifolia*), and fresh crushed garlic (*Allium sativum*) exhibited significant inhibitory effects against the tested bacterial strains. Tea Tree oil (*Melaleuca alternifolia*), and the crushed fresh garlic (*Allium sativum*) showed promising inhibitory activity even at low concentrations. In general, *E. coli, M. luteus* and *B. cereus* were the most susceptible. Therefore, the Tea Tree oil and crushed fresh garlic both showed significant antibacterial activity against the tested strains. The combination of tea tree oil and crushed fresh garlic exhibited a degree of antibacterial activity that was more than additive. Both tea tree oil and fresh crushed garlic separately and in combination may have potential for use in suppressing the growth of

pathogenic bacteria and could be used to develop a dose dependent practical application as antibacterial agents. Further research is warranted.

## References

Burt, S.A. (2004). Essential oils: their antibacterial properties and potential applications in foods: a review. *International Journal of Food Microbiology*. **94**: 223-253. 10.1016/j.ijfoodmicro.2004.03.022.

Gulsen, G. and Erol, A. (2010). Recent Patents on Anti-Infective Drug Discovery. 5: 91. https://doi.org/10.2174/157489110790112536

Prabuseenivasan, S., Jayakumar, M., & Ignacimuthu, S. (2006). *In vitro* antibacterial activity of some plant essential oils. *BMC Complementary and Alternative Medicine*, **6**, 39. http://doi.org/10.1186/1472-6882-6-39

Staveley, J. and Ramos, M. (2018). Antimicrobial Properties of Four Essential Oils. The Incubator Journal Frederick Community College. Vol.1, Issue 1.

Zaika, L. (1988). Spices and herbs: their antibacterial activity and its determination. *J Food Safety*, **23**:97-118.

## Bio

Dr. Silvia Godinez received her doctoral degree four years ago. She spent one year as a post-doc in Mexico, one year in Dr. Staveley's laboratory at Frederick Community College, and currently she is working in a second post-doctoral position at the CRAG in Barcelona, Spain.

# Contextual Label Smoothing with a Phylogenetic Tree on the iNaturalist 2018 Challenge Dataset

Michael J. Trammell[1], Priyanka Oberoi[1], James Egenrieder[2], John Kaufhold[1]

[1]General Dynamics Mission Systems' Deep Learning Analytics Center [2]Virginia Tech

## Abstract

Recognition of fine-grained visual categories (FGVC) in the natural world is a long-tailed problem, meaning recognizers must accurately recognize a large diversity of categories and most of those categories will naturally have limited training data, increasing the likelihood of overfitting in these many limited training data categories. The iNaturalist 2018 Challenge aimed to benchmark the state-of-the-art performance on species identification from a photo, where the long-tailed aspect of training is compounded by the visual similarity of many species. We demonstrate a new state of the art on the iNaturalist 2018 Challenge with Contextual Label Smoothing (CLS). CLS extends label smoothing to narrow the list of categories smoothed to only those within the same branch of a phylogenetic tree. CLS regularization improves performance significantly—the best publicly reported Top3 error reported on the iNaturalist 2018 Challenge was approximately 13%, which we improve to 12% with an ensemble of CLS networks trained with dynamic minibatching and additional inference windows. We present evidence that a 1% improvement on the FGVC iNaturalist 2018 Challenge test score (public score) represents over a 5 sigma improvement (test score stdev = 0.17 %) over the former state of the art.

## 1. Introduction

**THE PROBLEM OF FINE-GRAINED VISUAL CATEGORIZATION** (FGVC) has been studied across many domains with many image datasets, including FGVC-Aircraft [1], Stanford Cars [2], motorcycles [3] and shoes [4], among others. Many FGVC datasets of the natural world collect plant and animal species [5], birds [6], vegetables and fruits [7], plants [8], and dog breeds [9] to identify, among others. One of the largest and most imbalanced public datasets of natural imagery with these long-tailed FGVC challenges is the iNaturalist 2017 Challenge dataset, which the iNaturalist 2018 Challenge dataset made even larger and more imbalanced [10]. The iNaturalist 2018 Challenge training and validation data was made available by iNaturalist [11] and the competition was hosted on kaggle [12], which scored submissions on an unseen test set. Organizers of the iNaturalist 2018 Challenge aimed to:

> push the state of the art in automatic image classification for real world data that features a large number of fine-grained categories with high class imbalance. ... The dataset features many visually similar species, captured in a wide variety of situations, from all over the world. [12]

### 1.1. iNaturalist 2018's Long Tails

We call the most represented training categories in the iNaturalist 2018 Challenge data the "head" and the least represented categories the "tail" of the distribution (as in [13]). Recent work [13] has highlighted key properties of FGVC of long-tailed distributions: (1) there are many categories (2) most of the categories have limited training data (the tail categories) (3) error rates improve only when more labeled data is made available for the tail categories and (4) additional training data for the head categories does not appreciably improve overall performance (*i.e.* the network does not transfer learn from the head categories to the tail categories). On the iNaturalist 2018 Challenge data, approximately 10% of the categories (~800) comprise the head of the distribution, where each category has between 100 and 1000 training examples, and 75% of the categories (~6000) comprise the tail categories, where each category has between 2 and 30 training examples.

The prohibitive cost curve associated with generating sufficient training data for long-tailed FGVC applications to reach a threshold accuracy is sketched in [13]:

Collecting the eBird dataset took a few thousand motivated birders about 1 year. Increasing its size to the point that its top 2000 species contained at least $10^4$ images would take 100 years.

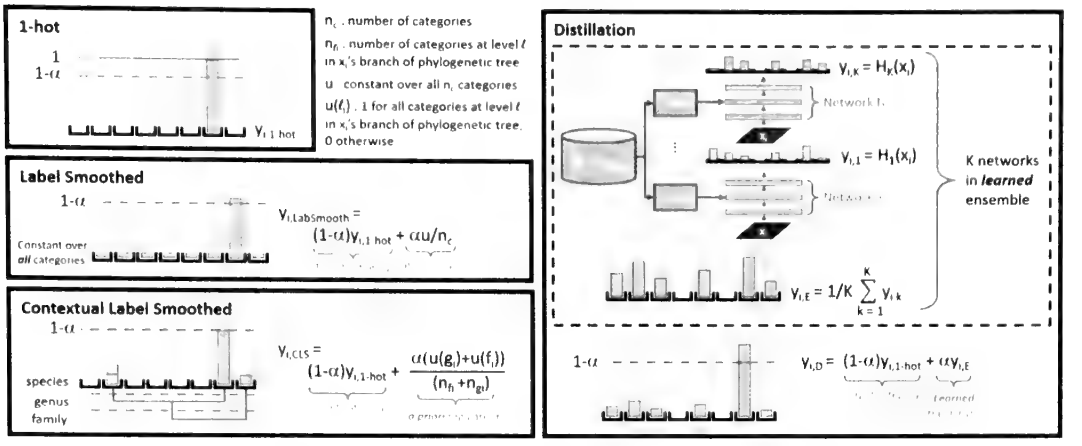



**Figure 1: Contextual Label Smoothing (CLS) label form compared to related label smoothing forms**: 1-hot encodings are *sparse labels* (top left). For example, for $x_i$, only one nonzero value in $y_{i,1\text{-hot}}$ is the target category and all others are 0s. 1-hot labels incorporate no regularization (either via a prior or learned post hoc from ensembling). Label smoothing (middle left), contextual label smoothing (bottom left), and distillation (right) all incorporate into their *full label vectors* some degree of regularization. In label smoothing, the regularizer is very weak but effective—$y_{i,\text{LabSmooth}}$ spreads out a small constant residual contribution of $\alpha/n_c$ to every category (where $n_c$ is the number of categories and u is a constant over all categories). In distillation, K classifiers are first trained with the 1-hot labels—the temperature-relaxed logits from the output layers of these K classifiers are then combined into a learned regularization term that is scaled and added to the 1-hot target category to form $y_{i,D}$. The distilled version's regularized $y_{i,D}$ has dense structure reflecting similarities among categories learned from the ensemble. Our method, contextual label smoothing (CLS), requires no learning as distillation does, and encodes label similarity from a phylogenetic tree into $y_{i,CLS}$. The number of categories shared at the genus and family level are $n_g$ and $n_f$, respectively. The notation $u(\ell_i)$ takes the value 1 for all categories shared at the $\ell$ level with the target category for $x_i$.

## 1.2. Label-efficient Approaches to Long Tails

For this reason, we seek more label-efficient approaches that incorporate context to address long-tailed FGVC challenges. Our aim is to efficiently encode in the labels, themselves, information that mitigates the performance degradation to tail categories stemming from limited training data. In the spirit of [14], in our proposed Contextual Label Smoothing (CLS), we allow tail categories to learn from training data pooled from similar categories as defined on a hierarchy (a phylogenetic tree) with label vector encodings (*i.e.* soft targets). This judicious form of label smoothing encodes information about which other categories are (likely to be) most similar, but unlike [14], we do not *learn* these relationships (which incurs a computational cost), but encode them directly with a portion of the phylogenetic tree [15] as the prior. The labels in the CLS approach are

diagrammed in contradistinction to 1-hot encoding, label smoothing and distillation.

While 1-hot label encodings (where one category is assigned a 1 and all others 0s) of categories have become common in mainstream object recognition [16]–[18], we argue these 1-hot independent category labels are label-inefficient—they do not effectively share informative training examples across similar labels; they are also overconfident—they make deep networks more susceptible to overfitting, especially on categories with limited training data.

Two simple relaxations of the 1-hot label encoding to better calibrate confidences in FGVC have been shown to improve (A) the robustness of the learned networks [19] and (B) the ability to learn more accurate tail categories post hoc from ensembles with limited training data [20]. In both label smoothing and distillation, the training labels are not 1-hot, but full, and retain some nonzero dot product from label vector to label vector. Inspired by both label smoothing and distillation, we demonstrate that contextual label smoothing (CLS), like hierarchical semantic encoding (HSE), can improve recognition rates on long-tailed FGVC problems.

### 1.3. CLS is Hierarchical Label Smoothing

Uniform label smoothing is an a priori decision to spread contributions from a target label over all other labels *uniformly*, which has the effect of penalizing overconfident predictions [19]. Intuitively, label smoothing allows *all* other categories to contribute training data to a target category, and spreading over *all* categories may spread the label information too thinly to efficiently transfer learn (as observed in [13]). In this work, we extend label smoothing to spread contributions from a label only within a branch of a phylogenetic tree provided a priori, not smooth over all other categories. Briefly, CLS exploits the phylogenetic tree to be more judicious about the label smoothing prior. Practically, we do not label smooth a training example of a humpback whale to have a nonzero contribution to learning a monarch butterfly category, but we do label smooth a training example of a gluphisia moth to have a nonzero contribution to learning the monarch butterfly category. While branches of phylogenetic trees are not always indicative of visual similarity, we empirically demonstrate that enough are to justify use of this prior.

### 1.4. CLS is Distillation with a Prior

Where distillation is an *empirical* post hoc approach to encode similarity into label vectors [20], our CLS work can be viewed as a form of

*a priori* distillation (Figure 2). Specifically, in distillation, an ensemble of classifiers are trained (from 1-hot labels). After learning, the (temperature-relaxed) logits of this ensemble empirically develop higher values for both the true category and visually similar categories. These post-hoc logits from this ensemble are added to the true 1-hot (hard targets) label for every training example in a downstream distillation of the ensemble. Intuitively, if only a handful of other classes are visually similar to the true class, when downstream training occurs with these distilled label vectors (soft targets), every one of those visually similar categories will contribute non negligibly to the training set for the original 1-hot target label. In this way, distillation *reuses* training examples from *other* categories to train to recognize the target categories most visually similar to it—this makes distillation a more label-efficient strategy than 1-hot encoding (Figure 2). CLS is an *a priori* version of distillation, encoding similarity as shared parentage on a phylogenetic tree provided without any downstream ensemble training (as are *learned* in either distillation or HSE).

### 1.5. Fine-Tuning with more Balanced Categories

On similar FGVC tasks [21], better performance was obtained by further fine-tuning on a more balanced subset of FGVC validation data with a small learning rate. Improvements on head categories with $\geq$100 training images were relatively small compared to tail categories with <100 training images. This provides an empirical rationale for fine-tuning on validation data more uniformly distributed over categories to improve performance on underrepresented tail categories. We incorporate this type of fine-tuning into CLS.

### 1.6. Contributions

We make a number of original contributions in this work:

- **Contribution 1: New State-of-the-Art on the iNaturalist 2018 Challenge.** We demonstrate a new state of the art result on the long-tailed FGVC iNaturalist 2018 Challenge Data [11]. We estimate through a prediction set that this new state-of-the-art outperforms the prior state-of-the-art by greater than 5 $\sigma$ on the unseen test data. We estimate the confidence interval of the score estimator for the unseen test data empirically via a Monte Carlo method. Specifically, we estimate the best fit line to the score computed by kaggle on the unseen test labels as a function of the score on the test score prediction set labels we do see to estimate the standard deviation of the estimator (see Figure 7 and Section 5 Test Score Prediction Analysis for

details).

- **Contribution 2: CLS works best with uniform sampling over categories.** In contradistinction to natural sampling advocated in [13], CLS benefits from uniform sampling of categories in training.
- **Contribution 3: CLS improves ensemble performance more per marginal network than other methods.** Given a choice between adding a network trained with some other technique to increase model diversity in an ensemble, adding another CLS-trained network is a better choice. This clarity can reduce the significant hyperparameter search and tuning costs over an ensemble.
- **Contribution 4: Larger Input Images Improve Performance.** While this is not a novel claim, we confirm empirically that larger input size images, which have recently been shown to improve performance on the same task without CLS [21], also improves performance of CLS.

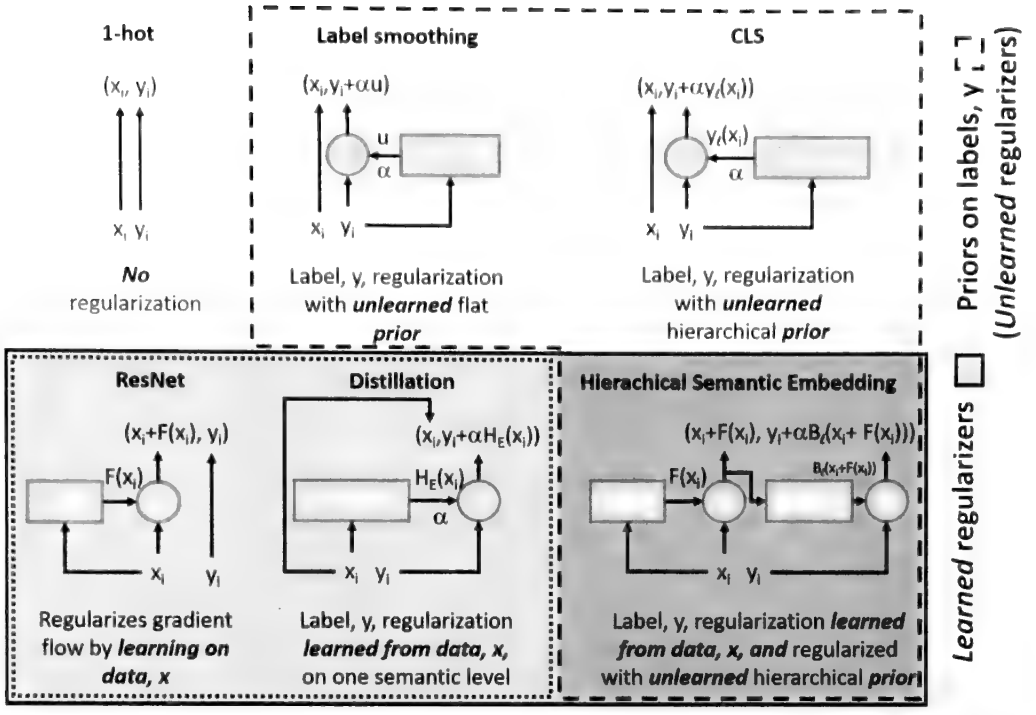



**Figure 2: Residual connection blocks regularize data and labels:** Five deep learning-based conceptual regularizer "blocks" to remedy overfitting and vanishing/noisy gradient issues of 1-hot label encodings (top left) are diagrammed. Across the top row are methods that only incorporate *unlearned* regularizers (*i.e.* priors only). Across the bottom row are methods which incorporate *learned* regularizers. On the bottom right, HSE incorporates both *learned and unlearned* regularizers. The well-known ResNet architecture ([22] bottom left) adds copies of the data, x, to *regularize gradients*—this architectural change is common to many of the other methods (both the trunk and branch networks of HSE [14] implement ResNet models, e.g.). Label smoothing ([19], top middle) can be viewed as a residual connection between a 1-hot $y_i$, and an *unlearned uniform* prior. This same strategy inspires this work on CLS (top right), but we use an *unlearned hierarchical* prior in the form of a phylogenetic tree. Distillation (bottom middle) can be viewed as a residual connection between a 1-hot $y_i$ and a *learned* soft target (the posterior distribution from learning an ensemble was used in [20]). The most general form of these combinations we have found is the very recent work on HSE (bottom right), which incorporates residual connections *learned* within trunk and branch networks, *learns* to update soft target priors based on an *unlearned* hierarchical prior, and combines these with residual connections at each level of the hierarchy.

## 2. Related Work

### 2.1. Deep Learning from 1-hot Labels

Since 2012 [17], deep networks have dominated the state of the art in object recognition on images, maturing year over year to include new network architectures [18], [22] until the performance of deep networks was on par with or better than human performance on a standard benchmark [23]. While significant attention has been paid to data augmentation [17], transfer learning [24], and new architectures [18], [22], less work has been

devoted to improving the 1-hot labels [19], [20], themselves, for training data. This work addresses improvements to the design of labels, themselves.

## 2.2. Label Vector Benefits

Work on improved label vector engineering includes label smoothing [19] and distillation [20], among others (Figure 2). Label smoothing is a simple method that incorporates a prior to drive deep networks to solutions with higher posterior entropy. Distillation, while originally proposed as a method to make networks smaller (in memory and computational cost of inference), has also demonstrated regularization and adversarial example defense properties.

Work on Hierarchical Semantic Embedding is most similar in spirit to this work, but achieves its goals of incorporating category similarity through a trunk and branches architecture over a collection of 1-hot label vectors at various semantic levels (from coarse to fine) [14]. Similar to distillation, it adds a predicted category score vector (*i.e.* a soft target) from a coarser level to the 1-hot label vector at the next finer level. FGVC results on three natural datasets, CUB [6], butterflies [14], and VegFru [7], demonstrate the value of HSE. HSE outperforms 17 other state of the art methods on CUB. The strategies employed in HSE appear to be more general than the simpler unlearned CLS prior proposed here (Figure 2), but HSE benefits have not yet been demonstrated on as large a dataset as that of the iNaturalist 2018 Challenge, which has >25x more fine-grained categories and >100x larger category imbalance, which are critically relevant aspects of long-tailed FGVC challenges [13].

Importantly, none of the datasets used to demonstrate HSE has more than 292 fine grained categories (compared to 8,142 for the iNaturalist Challenge 2018 data), with CUB's 200 categories separated into 122 genera, 37 families, and 13 orders, where 75% of CUB categories fall into the head category with 60 training images/category, and where all categories have at least 41 training images, for a max class imbalance of 1.5 (compared to 500 for the iNaturalist 2018 Challenge). The authors' new butterfly dataset also only contains 200 categories. This smaller scale of the FGVC challenges addressed by nascent exploration of HSE is encouraging, but qualitatively smaller scope than evaluation on iNaturalist Challenge 2018 data, which is an open dataset and more comprehensive than those datasets HSE authors chose to evaluate on.

Interestingly, HSE training develops learned attention mechanisms,

making a convincing case that without specifically labeled parts, HSE can learn features that exploit part-based attention to discriminate in FGVC, as was demonstrated to be critical for natural FGVC in other work [25]. The critical difference between the label vectors in HSE and our CLS work is that all of our label hierarchy information is encoded in label vectors without branches. CLS is a de facto *flat* prior that is *not learned* and is modularly separable from the architecture—*i.e.* there is only one label vector for each example in CLS, whereas HSE requires different label vectors at different levels in the architecture, increasing hyperparameter search costs.

### 2.3. Long-tailed FGVC Implications

The properties and implications of long-tailed distributions in FGVC have been summarized with convincing evidence [13] that (1) statistics of natural image categories are long-tailed, (2) more training data for head categories does not improve performance on tail categories, and (3) natural sampling of categories in training minibatches outperforms uniform sampling over categories. In [13], authors used standard 1-hot label encodings and sampled "naturally" (as opposed to uniformly) during training. The argument for natural over uniform sampling was empirical—results demonstrated both head and tail category performances both improved more with natural sampling. In contrast, we argue that the thoughtful vector encoding of labels with CLS overturns that guidance on sampling method (Contribution 2). Choosing training minibatches from CLS with uniform sampling over categories outperforms natural sampling. Authors conclude: "As a community we need to face up to the long-tailed challenge and start developing algorithms for image collections that mirror real-world statistics" which outlines the core motivation for this work [13].

### 2.4. Prior State of the Art iNaturalist Performance

The iNaturalist 2017 Challenge was won by Google (*GMV*, for Google Mountain View, on the leaderboard) with a Top5 error rate of less than 5% with an ensemble of InceptionV3 and InceptionV4 models trained at both 299x299 and 560x560 input image sizes, and subsequently fine-tuned on a balanced subset of the data left out of the test set [21]. The fine-tuning on balanced data boosts performance on tail categories of the dataset [1] and during inference 12 crops outperformed inference on a single prediction for the entire image.

Compared to the iNaturalist 2017 Challenge, the iNaturalist 2018 Challenge reduced the number of training images provided from 675,170 to

461,939, increased the number of classes from 5,089 to 8,142, and perhaps most significantly, provided a complete taxonomy for each class. A team from Dalian University won the 2018 challenge with a Top3 error rate of 13% [12]. Their winning ensemble consisted of 12 ResNet-152 models trained at both 320x320 and 392x392 input image sizes, six of which used matrix power normalized covariance pooling of the last layer of convolutional features [2].

## 3. Training Methodology

### 3.1. Training and Validation Data Set Splits

The iNaturalist 2018 Challenge data includes three mutually exclusive data sets: training, validation, and test data, each containing photos drawn from one of 8,142 species categories distributed over 4412 genera. The training data distribution is imbalanced, with the most represented species, *Branta canadensis* the "Canada goose", having 1000 training examples, whereas the least represented species in the training data is the *Spatula clypeata,* the "Northern shoveler duck," with only two training examples. The validation set is uniformly distributed over species, with three validation images per species. The test set labels are not provided to entrants, but entrants can submit Top3 label lists for each of the 149k test images to be scored on a Top3 error rate that is blind to which examples were marked correctly or incorrectly. In the development that follows, 2/3 of the validation data (two photos per species) is used for validation fine-tuning and 1/3 of the validation data (one photo per species) is used as the test score prediction set. In "vanilla" label smoothing [19], we assign the target label 0.8 and distribute the remaining 0.2 of that example to all other 8,141 categories in the label vector.

### 3.2. Initialization with Pretrained Networks

Closely following the winning *GMV* entrant from the iNaturalist 2017 Challenge, we start from an IRV2 and IV4 pretrained on ImageNet [18], [22]. These two network architectures are the starting points for training across all input sizes (299x299 and 598x598) and label smoothing methods (1-hot, vanilla label smoothing, and CLS). As in *GMV*, for each network in an ensemble, we strip the final layer of ImageNet-1K classes from the pretrained network and replace it with the iNaturalist 2017 output layer of 5,089 categories and sample minibatches of 32 images per minibatch without replacement from all training examples (we trained on 4 GPUs in parallel for an effective minibatch size of 128 for the IRV2 model and 6 GPUs in parallel for an effective minibatch size of 192 for the IV4

model). We fine-tuned on the iNaturalist 2017 training data for {80, 84} epochs for {IRV2, IV4}. We then fine-tuned on 90% of the iNaturalist 2017 validation data for {30, 14} epochs for {IRV2, IV4} using {8, 4} GPUs for effective minibatch sizes of {256, 128}. We used SGD with an initial learning rate of 0.018 and momentum=0.9 in the first round of training for the IRV2 model, reducing the learning rate by 10% every {8,6} epochs for {IRV2, IV4}. We used RMSProp for all other training. The second round of training began with learning rates of 0.002 for the IRV2 model and 0.001 for the IV4 model, and the training rate was multiplied by 0.9 every 10 epochs. Note that all minibatches in this pretraining were sampled naturally (as opposed to uniformly with replacement).

### 3.3. Base Fine-Tuning on iNaturalist 2018 Challenge Data

We strip the final layer of iNaturalist Challenge 2017 categories from each pretrained network and replace it with the iNaturalist 2018 Challenge output layer with 8,142 categories. When training, we sample minibatches uniformly over categories with replacement (*i.e.* we sample *uniformly*); this produces minibatches with approximately equal contributions from all 8,142 categories. We train for 1M-1.4M iterations using RMSprop with a base learning rate of 0.0045 in base fine-tuning. We use a batch size of 32. We retain only the model with the highest performance on the validation set, as assessed every 50k iterations.

### 3.4. Validation Fine-Tuning on iNaturalist 2018 Challenge Data

We fine-tune on the validation fine-tuning set only. The validation fine-tuning regime is identical to the base fine-tuning regime with the exception that training begins with a base learning rate of 0.0002, and continues for only 25k iterations.

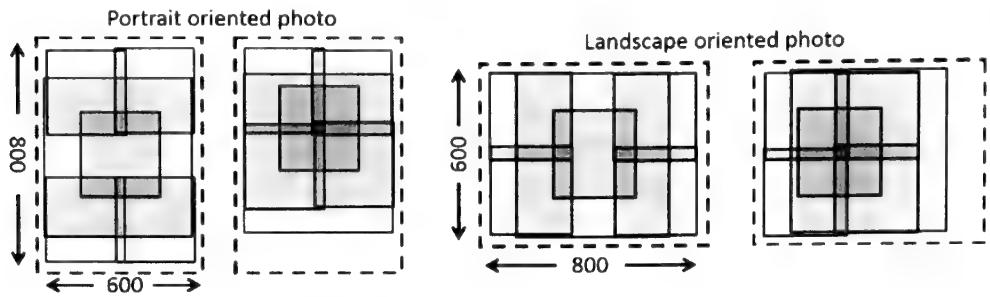



**Figure 3: Additional inference windows on a photo.** The "standard" twelve inference windows (six with the original image, the same six with the image flipped horizontally) are shown on the left of each orientation. For portrait-oriented photos, a second set of inferences is made on twelve more windows biased toward the top of the photo; for landscape-oriented photos, the second set of inferences is made on twelve more windows biased toward the left of the photo.

### 3.5. Ensembling

We compute unweighted model average ensemble results from multiple label smoothing methods to conduct a post hoc ablation study via ensemble composition. We rank the performance boosts from different components of the ensembles to assess the benefits of individual components of each ensemble. Ensemble components vary in input image size, network type, and label smoothing type.

### 3.6. Test Performance Error Analysis

**Additional inference windows:** When scoring, we include the standard middle, whole image, and four corner inference windows (with LR reflections). As an approximation to attention, we also include additional inference windows favoring the sides and top of the image calculated based on the aspect ratio of each image, under the assumption that this is where photographers are more likely to include the subject of the photo.

**Test score prediction error rates:** Nominally small (<0.5%) differences in Top3 error rates on leaderboards can be difficult to assess the relative merits of. By estimating a test score from the test score prediction data on many model outputs, we estimate a practical error bar on our test performances.

## 4. Results

The results collected here represent approximately 20,000 total GPU hours across a mix of NVIDIA GTX® 1080s, V100s and Titan® Xs.

For practical perspective, training a single one of our models through to final scoring on 2 GPUs requires approximately 10 days of compute on 299x299 input image sizes and 20 days on 598x598 input image sizes. Note that due to the size of our images and batches, only V100s can be used to train some of our models at our largest image sizes.

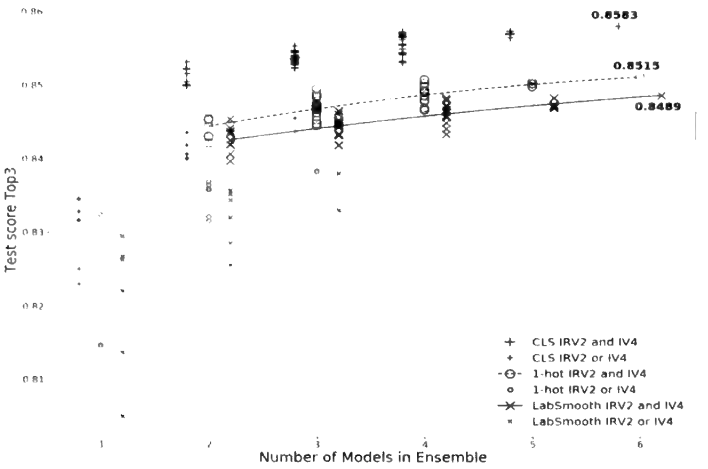



**Figure 4: CLS vs. label smoothing vs. 1-hot encodings.** CLS networks and ensembles of CLS networks outperform label smoothing and no label smoothing for both IRV2 and IV4 architectures assessed. The iNaturalist 2018 Challenge test scores returned from kaggle for the unseen test set is plotted vs. the number of models ensembled for each label smoothing method. A second-degree spline fit is plotted through the mean score of each set of IRV2 and IV4 ensembles for visual clarity.

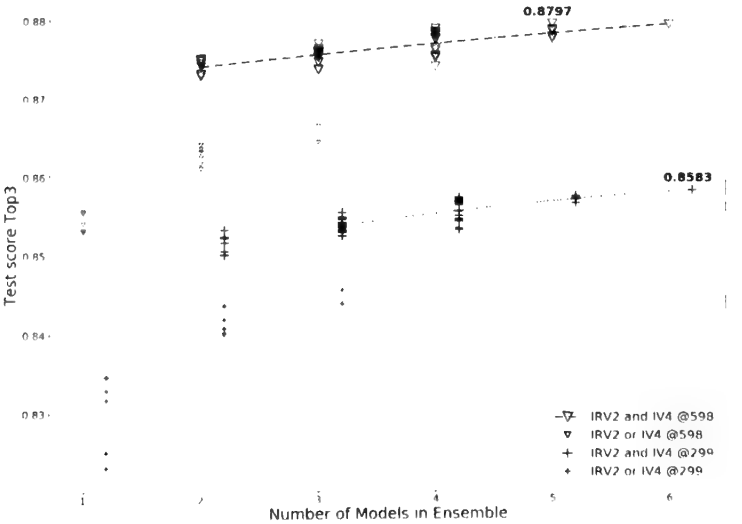



**Figure 5: CLS input size comparison.** We find that CLS on larger input image sizes (598x598) consistently outperforms CLS on smaller input image sizes (299x299). A second-degree spline fit is plotted through the mean score of each set of IRV2 and IV4 ensembles for visual clarity.

### 4.1. Label Smoothing Method Comparison

We show final iNaturalist 2018 Challenge test score results from kaggle on 299x299 pixel resolution images for the three label smoothing methods: 1-hot (*i.e.* no label smoothing), vanilla label smoothing (with 0.2 redistributed across all non-target classes), and CLS (with 0.2 redistributed across non-target classes in the same branch of the phylogenetic tree). Results of 3 runs each of {IRV2,IV4} and their ensembles demonstrate CLS outperforms both label smoothing and no label smoothing (*i.e.* 1-hot) encodings (Figure 4).

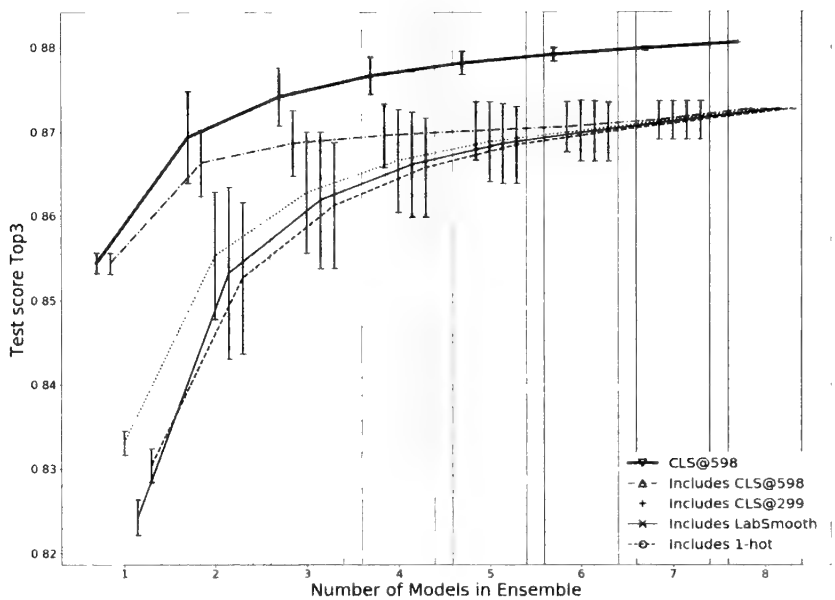



**Figure 6: Ensemble Ablation:** Only including 598x598 CLS networks in an ensemble with many networks provides state of the art performance with significantly reduced training and hyperparameter search and tuning costs compared to training a larger ensemble with a diversity of networks. Combining CLS networks trained with smaller input image sizes or networks not trained with CLS does not improve performance per network as much as adding another 598x598 CLS network (top curve).

### 4.2. Image Size Ensemble Ablation

We trained ensembles of CLS on both smaller (299x299) and larger (598x598) image input sizes into both IRV2 and IV4. The CLS performance on larger images consistently outperforms CLS trained on smaller images, whether on specific network types or ensembles of the same or different network types (Figure 5).

### 4.3. CLS Ensemble Ablation

Throughout testing, we find that additional CLS networks trained on larger input image sizes (598x598) improve ensembled results the most per

additional network in the ensemble. We find that unweighted network type diversity (including networks trained with and without label-smoothing, *i.e.* 1-hot, IRV2 and IV4 architectures, and smaller input image sizes) do not improve ensemble performance per additional network as much as adding a CLS-trained network at a 598x598 input image size, indicating that CLS with large imagery dominates the potential expected benefit of model diversity in these ensembles. When ensembles contain four or more networks, we observe that adding networks trained with either 1-hot or vanilla label smoothing label vectors can hurt performance.

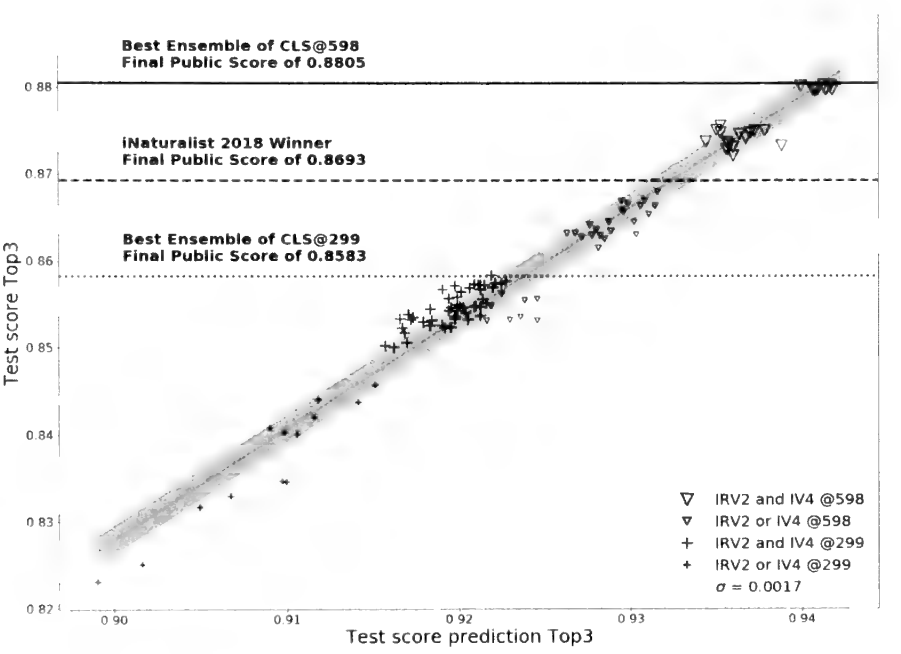



**Figure 7: Test Score Error Analysis:** By predicting the test error rate on the unseen test data based on a test score prediction subset (1/3) of the validation data we can observe, we develop confidence +/- 1-σ and 2-σ band estimates on the Test scores returned by the kaggle server on the unseen test data. The iNaturalist 2018 Challenge final Test score winner as reported on the iNaturalist 2018 Challenge leaderboard [12] at 13% Top3 error is shown as a dashed line.

## 4.4. Test Performance Error Analysis

Using an empirical Monte Carlo approach we develop a Test score predictor by fitting a line to the Test score as a function of the Test score prediction and from this we estimate that our new CLS state of the art result on iNaturalist 2018 Challenge test score has a +/- 0.17% σ error (

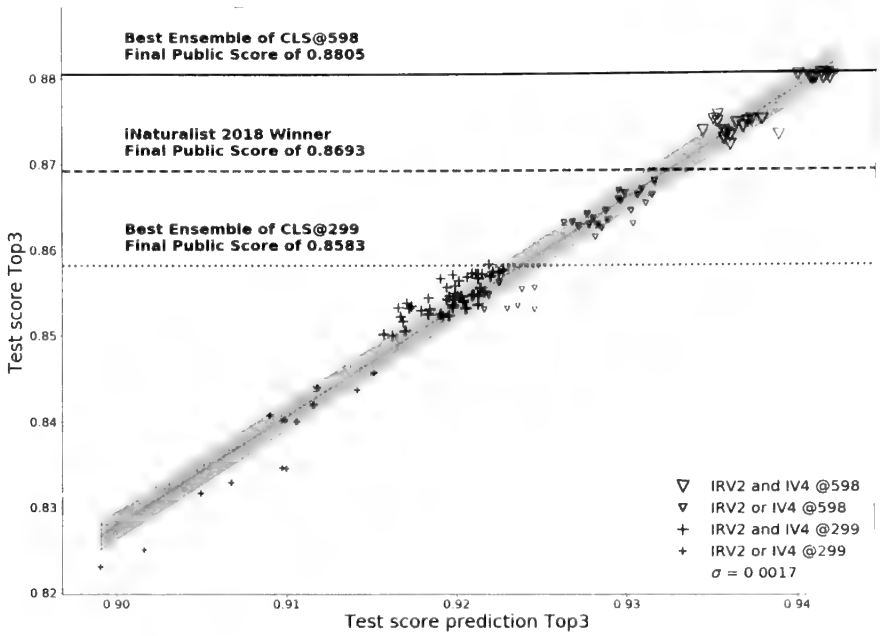



**Figure 7**). Our 1.0% improvement over the former state of the art represents a greater than 5 σ improvement over the best prior reported public test score of 0.8693 (compared to our 0.8805) with this estimate of score variability.

## 5. Discussion

**CLS shares training data among categories:** By encoding non-zero values, representative of proximity in the phylogenetic tree, in the label vectors for categories that are not the true target category, CLS learns from a more diverse set of examples than only those formally labeled as the putative target type. In long-tailed FGVC tasks, we expect a number of benefits from this approach.

In theory, for each target tail category, the relatively few training examples of that category with their much larger label vector component (0.8) will anchor the learned latent space of activations for that category with data from that target category. Without full vector labels of any type (*i.e.* 1-hot labels), the deep network could overfit to these relatively few training examples of the target category (*i.e.* memorize them), suffering poor generalization with no other information available to prevent this overfitting. Relatively fewer categories (but each with more training examples) from the head of the distribution that share the same branch of the phylogenetic tree as the target category will also contribute to training the target category. These examples will bias the learned latent space of activations for the target tail category to move closer to those related head

categories, encouraging transfer learning from the head to the tail. Relatively more non-target tail categories, each with fewer examples, will more diffusely contribute to training the target tail category, ensuring that the network does not overfit to either the relatively fewer training examples of the target tail category or the more represented contributing head categories.

In practice, any of these three effects may dominate, and rigorously calibrating them is left for future work devoted to that detailed analysis to compare to HSE. In addition to the rich relationships we exploit to improve discriminative performance of species identification, it is also possible that this approach could inform related research on ontological views of relationships between different species. Specifically, the data-rich categories from the head of the distribution might be used to stabilize, communicate, and/or extend categorical relationships across hierarchies (including predicates on the taxonomic relationships).

**Focused Ensemble Performance with One Label Smoothing Method:** Since each CLS network at the 598x598 input size added to an ensemble improves performance more than adding another marginal network, this CLS benefit also reduces training time by focusing only on the CLS-trained models. For instance, in our ensemble ablation, we see that five CLS networks trained at the 598x598 input image size outperforms five CLS networks with the addition of any other network type that is not CLS 598x598. This clarity allows us to focus computational resources on only one type of network and not risk losing potentially beneficial diversity in our ensembles that might accrue from other models with complementary strengths had we trained them. This is a critical benefit to downstream work comparing different methods because it guides efficient allocation of limited compute resources on an already computationally intensive task.

**Test Score Prediction Analysis:** The scores from the test score prediction set (part of the validation set, which entrants see) are highly correlated with the test scores for the same model (network, or ensemble of networks, e.g) on the unseen test data provided per blinded submission by kaggle (

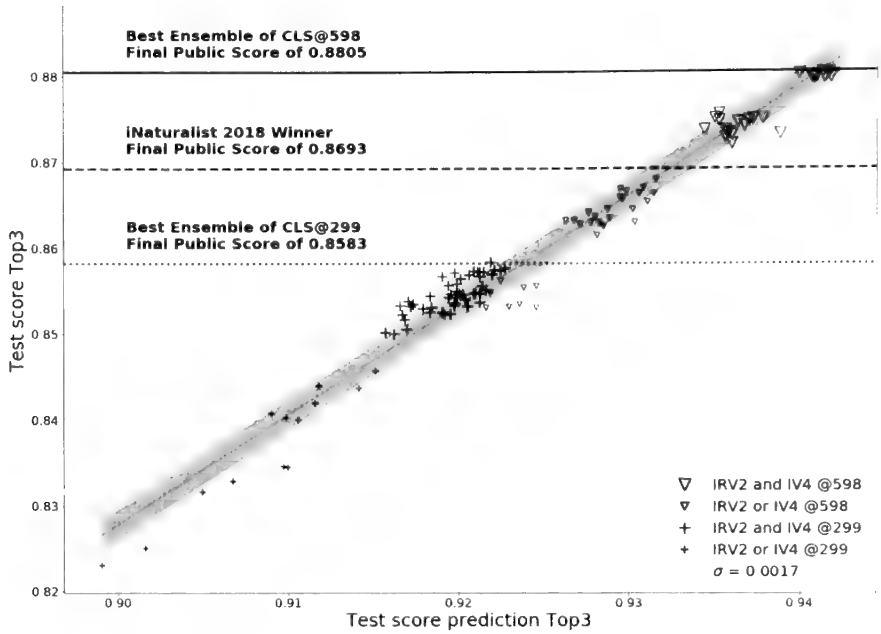



**Figure 7**). In independent testing, we submitted a number of single category labels to kaggle to interrogate the iNaturalist 2018 Challenge test data and found in each case that the resulting test scores were very close to each other. This indicated that the mutually exclusive test set, while unseen and held out from training and validation data, was likely uniformly distributed over categories, as was the provided validation set. Based on this insight, we used only a portion of the validation set for validation fine-tuning (following [21]), leaving out a portion also uniformly distributed over categories to predict the Test score. We found that a score computed on this Test score prediction set was highly correlated with the actual Test score.

We note the interrogation of the test set in this way does not confer significant benefit on the Test score, as relatively tight bounds can be estimated [25], and that large numbers of submissions will typically not improve test scores. To wit, we did not tune, nor overfit to the test set here, except to establish that it was uniformly distributed over categories.

By predicting the Test score from a presumably identically distributed (over categories) Test score prediction set, we estimate a conservative error bar on the Test score—meaning that the actual error bar is likely smaller than our estimate. Specifically, the error bar fit estimate degrades with both the Test score variability on the y-axis (the iNaturalist 2018 Challenge test score $\sigma$ we seek to estimate) as well as the prediction test set score variability on the x-axis (which is a nuisance parameter). We

cannot separate out these two sources of variability, but since the test set has many more examples in it, we anticipate its contribution to the estimation error, $\sigma_{test}$, is smaller than the contribution to the estimation error of the Test score prediction set, $\sigma_{predict}$.

This error analysis helps in two ways. First, it provides a rough measure of the real performance improvement from method to method based on an empirically estimated confidence interval. Roughly, for CLS that translates to slightly larger than an approximately 5 $\sigma$ improvement over the former state-of-the-art reported on the iNaturalist 2018 Challenge [12]. Second, and more important to guide future work, such an estimation error together with the measured performance improvement per marginal ensemble network provides a rough means to estimate the expected performance improvement per additional trained network in an ensemble. This provides an ensembling stopping criterion to focus compute resources, which, along with the insight of Contribution 3, that CLS improves ensemble performance more per marginal network than other methods, is critical to efficiently allocating compute resources for methodological comparisons at scale (such as between CLS and HSE, e.g.) in downstream work.

**Improving Tail Category Performance with Fine-Tuning:** Prior work [21] inspired our adoption of fine-tuning on a more uniformly distributed set of categories. In our case, we used a fraction of the validation data for this purpose. We see similar gains in this work—*i.e.* CLS also benefits from this fine-tuning approach.

## 6. Conclusion

The long tails of FGVC tasks for natural image corpora present daunting training data collection requirements to achieve required accuracy objectives on tail categories with mainstream deep learning methods. Namely, the tail categories are many, sparse, and similar, making their per-category accuracies difficult to improve on with 1-hot labels that treat them independently in training. In this work we demonstrate that CLS' hierarchical prior on vector labels in the form of a phylogenetic tree can pool training data contributions from many of the tail classes, exploit their similarities, and thereby improve the accuracy on tail classes compared to 1-hot labels or other less judicious vector label smoothings.

**CLS is Encoded by Domain Experts:** The benefit of CLS alone is significant and does not require expertise in deep learning to realize—the phylogenetic tree prior came directly from a phylogenetic tree curated by

biologists [15]. This is the only change from other methods [21] benchmarked on this same dataset that we show underperform without CLS compared to the same methods incorporating CLS.

**CLS is Compatible with more Data-Driven Methods:** While we present results only on CLS without a CLS-specific hyperparameter search, the CLS method proposed is compatible with more empirical distillation and HSE methods which adjust label vectors based on training. Specifically, CLS can be incorporated directly into the trunk network of HSE, for instance. The CLS ensembles can be distilled into a single network to realize the benefits of distillation, including distillation benefits of adversarial example defense and compute reduction, *e.g.*

**CLS's Prior Models can be Extended by Human or Machine:** While we demonstrate a simple CLS approach that exploits an a priori provided phylogenetic tree, this *unlearned* prior can very likely be improved because the phylogenetic tree is not, by design, a guide to visual similarity, even within a species. For instance, even within species, there can be further training example pooling with visual similarity as encoded through latent activation clustering. Among butterflies, for instance, the within-species separation of chrysalis, caterpillar and butterfly stages may create separable clusters in an embedding of latent activations (as with t-SNE, e.g.). Within a bird species, the visual ornamentation of males vs. females may similarly cluster in an embedding of latent activations. Similarly, dog breeds may cluster. All of these finer levels may be similarly encoded into the CLS prior by either machine or human curator. As with all FGVC tasks, this presents additional challenges as training data fragments among the categories because categories with very little training data are split further, dividing the sparse training data among the finer subcategories. We show that CLS can still effectively pool training data in that scenario at the genus to species level of granularity and leave for future work the demonstration of even more fine-grained applications of CLS.

**Future Work:** Demonstrating and evaluating the combined benefits of both the *a priori* hierarchical CLS prior and the post hoc *learned* latent encodings of similarities (as in HSE and distillation, e.g.) together is left for future work, as is the significant challenge of comparing other methods that make use of the phylogenetic tree prior (like HSE) to CLS on the scale of the iNaturalist 2018 dataset. For perspective, even with no CLS hyperparameter tuning, the present study required >20,000 of GPU compute time. The GPU compute costs of rigorously comparing HSE to CLS with the hyperparameter searches required to reach conclusive results are

anticipated to be even larger, and may warrant additional AutoML investigations, further increasing the computational costs.

## 7. Acknowledgements

This research was developed with funding from the Defense Advanced Research Projects Agency. The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government. We thank the iNaturalist organization for providing the iNaturalist 2018 dataset, phylogenetic tree, and the held out test data for Kaggle blind scoring. We thank the anonymous reviewers for clarifying revisions and highlighting important themes we had not sufficiently emphasized in the original draft.

## References

[1] S. Maji, E. Rahtu, J. Kannala, M. Blaschko, and A. Vedaldi, "Fine-Grained Visual Classification of Aircraft," Jun. 2013.

[2] J. Krause, M. Stark, J. Deng, and L. Fei-Fei, "3d object representations for fine-grained categorization," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2013, pp. 554–561.

[3] A. B. Hillel and D. Weinshall, "Subordinate class recognition using relational object models," in *Advances in Neural Information Processing Systems*, 2007, pp. 73–80.

[4] T. L. Berg, A. C. Berg, and J. Shih, "Automatic attribute discovery and characterization from noisy web data," in *European Conference on Computer Vision*, 2010, pp. 663–676.

[5] G. Van Horn *et al.*, "The inaturalist species classification and detection dataset," 2018.

[6] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, "The caltech-ucsd birds-200-2011 dataset," 2011.

[7] S. Hou, Y. Feng, and Z. Wang, "Vegfru: A domain-specific dataset for fine-grained visual categorization," in *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017, pp. 541–549.

[8] H. Goeau, P. Bonnet, and A. Joly, "Plant identification based on noisy web data: the amazing performance of deep learning (LifeCLEF 2017)," in *CLEF 2017-Conference and Labs of the Evaluation Forum*, 2017, pp. 1–13.

[9] J. Liu, A. Kanazawa, D. Jacobs, and P. Belhumeur, "Dog breed classification using part localization," in *European Conference on Computer Vision*, 2012, pp. 172–185.
[10] "iNaturalist_Competition_FGVC_2018.pdf." [Online]. Available: https://www.dropbox.com/s/52nz6qc3zcwqhoa/iNaturalist_Competition_FGVC_2018.pdf?dl=0. [Accessed: 15-Nov-2018].
[11] *iNaturalist competition details. Contribute to visipedia/inat_comp development by creating an account on GitHub*. Visipedia, 2018.
[12] "iNaturalist Challenge at FGVC5." [Online]. Available: https://www.kaggle.com/c/inaturalist-2018. [Accessed: 01-Jul-2018].
[13] G. Van Horn and P. Perona, "The Devil is in the Tails: Fine-grained Classification in the Wild," *ArXiv Prepr. ArXiv170901450*, 2017.
[14] T. Chen, W. Wu, Y. Gao, L. Dong, X. Luo, and L. Lin, "Fine-grained representation learning and recognition by exploiting hierarchical semantic embedding," *ArXiv Prepr. ArXiv180804505*, 2018.
[15] "GBIF." [Online]. Available: https://www.gbif.org/. [Accessed: 15-Nov-2018].
[16] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning.," in *AAAI*, 2017, vol. 4, p. 12.
[17] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, 2012, pp. 1106–1114.
[18] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.
[19] G. Pereyra, G. Tucker, J. Chorowski, \Lukasz Kaiser, and G. Hinton, "Regularizing neural networks by penalizing confident output distributions," *ArXiv Prepr. ArXiv170106548*, 2017.
[20] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *ArXiv Prepr. ArXiv150302531*, 2015.
[21] Y. Cui, Y. Song, C. Sun, A. Howard, and S. Belongie, "Large Scale Fine-Grained Categorization and Domain-Specific Transfer Learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4109–4118.
[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *ArXiv Prepr. ArXiv151203385*, 2015.

[23] O. Russakovsky *et al.*, "Imagenet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.
[24] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," in *Advances in neural information processing systems*, 2014, pp. 3320–3328.
[25] K. Kawaguchi, L. P. Kaelbling, and Y. Bengio, "Generalization in deep learning," *ArXiv Prepr. ArXiv171005468*, 2017.

## Bios

**Michael Jeremy Trammell** is a software engineer at Deep Learning Analytics and leads the Deep BioThreatID project. He was recognized at IEEE's CVPR conference in 2018 as part of the team that placed 2nd in the world in the iNaturalist 2018 Challenge.

**Priyanka Oberoi** is a data scientist and head of ethics and fairness in machine learning at Deep Learning Analytics. She received her Masters degree in Biotechnology and Bioinformatics from Johns Hopkins University. She was recognized at IEEE's CVPR conference in 2018 as part of the team that placed 2nd in the world in the iNaturalist 2018 Challenge.

**Jim Egenrieder** is a fish and wildlife biologist and teaches Biodiversity Stewardship and Watershed Systems Stewardship at Virginia Tech's Center for Leadership in Global Sustainability in the National Capital Region. He is also on the Research Faculty of Virginia Tech's College of Engineering and is Director of the Virginia Tech Thinkabit Lab™ teaching engineering and programming of microcontroller and microprocessor circuits.

**John Kaufhold** is the Founder of Deep Learning Analytics, a machine learning startup in the DC Metro Region. He received his Ph.D. in Biomedical Engineering from Boston University as a Whitaker Fellow, was named a Technical Fellow of SAIC (Leidos), and was recently named a Fellow of the Washington Academy of Sciences. He was recognized at IEEE's CVPR conference in 2018 as part of the team that placed 2nd in the world in the iNaturalist 2018 Challenge.


Washington Academy of Sciences
1200 New York Avenue
Rm G119
Washington, DC 20005

Please fill in the blanks and send your application to the address above. We will contact you as soon as your application has been reviewed by the Membership Committee. Thank you for your interest in the Washington Academy of Sciences.

(Dr. Mrs. Mr. Ms)

_____

Business Address

_____

Home Address

_____

Email

_____

Phone _____

Cell Phone _____

preferred mailing address Type of membership

____Business ____Home ____Regular ____Student

| Schools of Higher Education attended | Degrees | Dates |
|---|---|---|
| | | |
| | | |
| | | |

Present Occupation or Professional Position _____
Please list memberships in scientific societies – include office held

_____

_____

_____

## Instructions to Authors

1. Deadlines for quarterly submissions are:

   Spring – February 1 Fall – August 1
   Summer – May 1 Winter – November 1

2. Draft Manuscripts using a word processing program (such as MSWord), not PDF. We do not accept PDF manuscripts.
3. Papers should be 6,000 words or fewer. If there are 7 or more graphics, reduce the number of words by 500 for each graphic.
4. Include an abstract of 150-200 words.
5. Include a two to three sentence bio of the authors.
6. Graphics must be in greytone, and be easily resizable by the editors to fit the Journal's page size. Reference the graphic in the text.
7. Use endnotes or footnotes. The bibliography may be in a style considered standard for the discipline or professional field represented by the paper.
8. Submit papers as email attachments to the editor or to wasjournal@washacadsci.org .
9. Include the author's name, affiliation, and contact information – including postal address. Membership in an Academy-affiliated society may also be noted. It is not required.
10. Manuscripts are peer reviewed and become the property of the Washington Academy of Sciences.
11. There are no page charges.
12. Manuscripts can be accepted by any of the Board of Discipline Editors.

# Washington Academy of Sciences
# Affiliated Institutions

National Institute for Standards & Technology (NIST)

Meadowlark Botanical Gardens

The John W. Kluge Center of the Library of Congress

Potomac Overlook Regional Park

Koshland Science Museum

American Registry of Pathology

Living Oceans Foundation

National Rural Electric Cooperative Association (NRECA)

# Delegates to the Washington Academy of Sciences Representing Affiliated Scientific Societies

| | |
|---|---|
| Acoustical Society of America | Paul Arveson |
| American/International Association of Dental Research | J. Terrell Hoffeld |
| American Assoc. of Physics Teachers, Chesapeake Section | Frank R. Haig, S. J. |
| American Astronomical Society | Sethanne Howard |
| American Fisheries Society | Lee Benaka |
| American Institute of Aeronautics and Astronautics | David W. Brandt |
| American Institute of Mining, Metallurgy & Exploration | E. Lee Bray |
| American Meteorological Society | Vacant |
| American Nuclear Society | Charles Martin |
| American Phytopathological Society | Vacant |
| American Society for Cybernetics | Stuart Umpleby |
| American Society for Microbiology | Vacant |
| American Society of Civil Engineers | Vacant |
| American Society of Mechanical Engineers | Daniel J. Vavrick |
| American Society of Plant Physiology | Mark Holland |
| Anthropological Society of Washington | Vacant |
| ASM International | Toni Marechaux |
| Association for Women in Science | Jodi Wesemann |
| Association for Computing Machinery | Vacant |
| Association for Science, Technology, and Innovation | F. Douglas Witherspoon |
| Association of Information Technology Professionals | Vacant |
| Biological Society of Washington | Vacant |
| Botanical Society of Washington | Chris Puttock |
| Capital Area Food Protection Association | Keith Lempel |
| Chemical Society of Washington | Vacant |
| District of Columbia Institute of Chemists | Vacant |
| Eastern Sociological Society | Ronald W. Mandersheid |
| Electrochemical Society | Vacant |
| Entomological Society of Washington | Vacant |
| Geological Society of Washington | Jurate Landwehr |
| Historical Society of Washington DC | Vacant |
| Human Factors and Ergonomics Society | Gerald Krueger |

(continued on next page)

# Delegates to the Washington Academy of Sciences Representing Affiliated Scientific Societies

(continued from previous page)

| | |
|---|---|
| Institute of Electrical and Electronics Engineers, Washington Section | Richard Hill |
| Institute of Food Technologies, Washington DC Section | Taylor Wallace |
| Institute of Industrial Engineers, National Capital Chapter | Neal F. Schmeidler |
| International Association for Dental Research, American Section | Christopher Fox |
| International Society for the Systems Sciences | Vacant |
| International Society of Automation, Baltimore Washington Section | Richard Sommerfield |
| Instrument Society of America | Hank Hegner |
| Marine Technology Society | Jake Sobin |
| Maryland Native Plant Society | Vacant |
| Mathematical Association of America, Maryland-District of Columbia-Virginia Section | John Hamman |
| Medical Society of the District of Columbia | Julian Craig |
| National Capital Area Skeptics | Vacant |
| National Capital Astronomers | Jay H. Miller |
| National Geographic Society | Vacant |
| Optical Society of America, National Capital Section | Jim Heaney |
| Pest Science Society of America | Vacant |
| Philosophical Society of Washington | Larry S. Millstein |
| Society for Experimental Biology and Medicine | Vacant |
| Society of American Foresters, National Capital Society | Marilyn Buford |
| Society of American Military Engineers, Washington DC Post | Vacant |
| Society of Manufacturing Engineers, Washington DC Chapter | Vacant |
| Society of Mining, Metallurgy, and Exploration, Inc., Washington DC Section | E. Lee Bray |
| Soil and Water Conservation Society, National Capital Chapter | Erika Larsen |
| Technology Transfer Society, Washington Area Chapter | Richard Leshuk |
| Virginia Native Plant Society, Potowmack Chapter | Alan Ford |
| Washington DC Chapter of the Institute for Operations Research and the Management Sciences (WINFORMS) | Meagan Pitluck-Schmitt |
| Washington Evolutionary Systems Society | Vacant |
| Washington History of Science Club | Albert G. Gluckman |
| Washington Paint Technology Group | Vacant |
| Washington Society of Engineers | Alvin Reiner |
| Washington Society for the History of Medicine | Alain Touwaide |
| Washington Statistical Society | Michael P. Cohen |
| World Future Society, National Capital Region Chapter | Jim Honig |

Washington Academy of Sciences
Room GL117
1200 New York Ave. NW
Washington, DC 20005
Return Postage Guaranteed